GEORGIA INSTITUTE OF TECHNOLOGY — OFFICE OF CONTRACT ADMINISTRATION

PROJECT ADMINISTRATION DATA SHEET

[X] ORIGINAL [ ] REVISION NO. ______

Project No. E-21-627 — DATE 5/4/82

Project Director: M. H. Hayes — School/~~XXX~~Lab Electrical Engineering

Sponsor: National Science Foundation, Washington, D. C. 20550

Type Agreement: Grant No. ECS-8204793

Award Period: From 7/1/82 To ~~12/31/83~~* 12/31/84 (Performance) ______ (Reports)

Sponsor Amount: $24,552 — Contracted through: GTRI/~~GIT~~

Cost Sharing: $7,524 (E-21-375)

Title: Research Initiation: Signal Reconstruction from Partial Phase and Magnitude Information

ADMINISTRATIVE DATA

OCA Contact William F. Brown X4820

1) Sponsor Technical Contact:

Elias Schutzman, Program Official
Division of Electrical, Computer and Systems Engineering
National Science Foundation
Washington, D.C. 20550
(202) 357-9618

2) Sponsor Admin/Contractual Matters:

Lee A. DeHerrera, Grants Official
Division of Grants & Contracts
Directorate for Administration
National Science Foundation
Washington, D.C. 20550
(202) 357-9602

Defense Priority Rating: None

Security Classification: None

RESTRICTIONS

See Attached NSF Supplemental Information Sheet for Additional Requirements.

Travel: Foreign travel must have prior approval — Contact OCA in each case. Domestic travel requires sponsor approval where total will exceed greater of $500 or 125% of approved proposal budget category.

Equipment: Title vests with GIT

COMMENTS:

* Includes the usual six-month unfunded flexibility period.

RECEIVED MAY 1982 Research Reports Office

COPIES TO:

Research Administrative Network

~~Administrative Coordinator~~	Research Security Services	EES Public Relations (2)
Research Property Management	~~Reports Coordinator (OCA)~~	Computer Input
Accounting	Legal Services (OCA)	Project File
Procurement/EES Supply Services	Library	Other ______

FORM OCA 4:781

5R519-2-N-8

GEORGIA INSTITUTE OF TECHNOLOGY OFFICE OF CONTRACT ADMINISTRATION

SPONSORED PROJECT TERMINATION/CLOSEOUT SHEET

Date 9/9/85

Project No. E-21-627 School/~~Lab~~ EE

Includes Subproject No.(s)

Project Director(s) M. H. Hayes GTRC / ~~GIT~~

Sponsor National Science Foundation

Title Research Initiation: Signal Reconstruction from Partial and Magnitude Information

Effective Completion Date: 12/31/84 (Performance) 3/31/85 (Reports)

Grant/Contract Closeout Actions Remaining:

- [ ] None
- [ ] Final Invoice or Final Fiscal Report
- [ ] Closing Documents
- [X] ~~Final Report of Inventions~~ Patent Questionnaire
- [ ] Govt. Property Inventory & Related Certificate
- [ ] Classified Material Certificate
- [ ] Other

Continues Project No. Continued by Project No.

COPIES TO:

Project Director
Research Administrative Network
Research Property Management
Accounting
Procurement/GTRI Supply Services
Research Security Services
Reports Coordinator (OCA)
Legal Services

Library
GTRC
Research Communications (2)
Project File
Other A. Jones; M. Heyser

FORM OCA 69.285

Signal Reconstruction From

Partial Phase and Magnitude Information

NSF Research Initiation Grant
Number ECS-8204793

Progress Report

and

Request for Continued Support

Monson H. Hayes
School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

Table of Contents

I. SUMMARY OF PROGRESS

1.1 Introduction

The research funded in this grant is concerned with the reconstruction of signals from partial phase and magnitude information. In particular, the practical issues involved in reconstructing a discrete time signal from the magnitude of its Fourier transform were to be studied. In addition, the problem of reconstructing a discrete-time signal from noisy phase measurements were to be studied. In the next two sections, a summary of the progress during the first year of this grant on these problems are presented. Following this summary, publications prepared during this first year are listed and a report on the graduate student involvement in this research is described.

1.2 Phase Retrieval

The phase retrieval problem is concerned with the reconstruction of a signal given only knowledge of the magnitude of its Fourier transform. For discrete-time signals, the uniqueness of the solution to the phase-retrieval problem has recently been considered in detail [4,5]. Specifically, it has been shown that for one-dimensional signals spectral magnitude is not sufficient to uniquely define the signal. For multi-dimensional signals, on the other hand, except for trivial ambiguities the uniqueness is quaranteed if the signal has a multidimensional z-transform which is an irreducible polynomial. In spite of this uniqueness, however, it is not generally possible to reconstruct a multidimensional signal from only its spectral magnitude. As part of the funded research, an investigation was to be undertaken to explain the observation that very little phase information seems to be required in order to reconstruct multidimensional signals from their spectral magnitude. This study has been very fruitful and has led to a better understanding of the theoretical issues pertaining to the phase and magnitude retrieval problems. Specifically, a new way of representing spectral information was established which involves writing the Fourier transform of a discrete time signal in terms of its spectral "amplitude" and spectral "angle" rather than its spectral magnitude and phase [2]. Just as with magnitude and phase, amplitude and angle are both necessary in order to uniquely define a complex number. Spectral angle information is equivalent to the knowledge of the tangent of the phase and spectral amplitude is equivalent to knowledge of the magnitude of the transform along with one bit of phase information. With such a representation of spectral information, it was shown theoretically that any causal finite length sequence is uniquely defined in terms of its spectral amplitude [1,2]. Specifically, Theorem: If $x(n)$ and $y(n)$ are two causal finite length sequences having the same spectral amplitude, then $x(n)=y(n)$.

With this result, a duality was established which resembles the duality found in the uniqueness of a minimum phase sequence in terms of its spectral magnitude or spectral phase. Specifically:

Duality Property: A sequence with a causal cepstrum is uniquely defined by the magnitude or to within a scale factor by the phase of its Fourier transform. Similarly, a causal (finite length) sequence is uniquely defined by the amplitude or to within a scale factor by the angle of its Fourier transform if $X(z)$ contains no conjugate reciprocal zeros.

Once the uniqueness of a finite-length causal signal in terms of its spectral amplitude was established, the definition of spectral was then generalized to include a larger class of signal reconstruction problems. Specifically, the one bit of phase associated with the spectral magnitude was extended to include any division of the complex plane into two half planes [1].

Finally, the issues involved in the reconstruction of a signal from its spectral amplitude and angle were addressed [1,2]. As has been previously demonstrated [4,5], reconstructing a signal from its spectral angle is straight-forward. The uniqueness of a signal in terms of its spectral amplitude, however, generally requires that the amplitude be known for all frequencies. Although only a finite number of amplitude samples are theoretically necessary, the specific frequencies for which the amplitude values are required cannot be determined a priori. Nevertheless, it has been experimentally observed that if the amplitude of the DFT of the signal is known and if the length of the DFT is large compared with the length of the signal then the solution was generally unique. Therefore, using an iterative approach similar to that introduced by Gerchberg and Saxton [6], three different iterative approaches were developed for reconstructing a signal from its amplitude. The difference between these algorithms lies in the method used to impose the known amplitude information in the iteration. In almost all cases, these algorithms successfully reconstructed the signal. However, a theoretical proof that the iteration will always converge to the correct signal (provided that the solution is uniquely defined by the given set of amplitude samples) has not yet been established.

1.3 Magnitude Retrieval

The magnitude retrieval problem is characterized by the desire to reconstruct a signal from only its spectral phase.

It has been previously demonstrated that most finite length signals are uniquely specified by the phase of their Fourier transforms and a number of algorithms have been developed for reconstructing these finite length signals from their spectral phase [4,5,7]. Unfortunately, however, it has been found that the reconstructed signals are particularly sensitive to noise. Therefore, due to its importance in applying phase-only reconstruction techniques to practical problems, part of the funded research has been directed towards an examination, both experimentally and analytically, into the reconstruction errors which result from phase measurement errors and computational noise.

In order to gain a better understanding of the issues involved, three "non-stochastic" algorithms were developed which attempt to make the signal reconstruction more robust in the presence of noise. These algorithms have been termed "non-stochastic" since they do not take into account any information or assumptions about the statistics of the phase noise. In all cases the underlying assumption was that noise corrupted measurements of the phase of a finite length sequence of length N were specified at kN distinct frequencies in the interval $(0,\pi)$. Since only N samples are required for phase-only signal reconstruction in the noise free case, redundancy is provided by the excess number of amplitude samples. The first method which was investigated performed k reconstructions of x(n) from N phase samples and then averaged the reconstucted sequences [8]. The second method considered the possibility of smoothing or averaging the phase samples in order to obtain better phase estimates to use in the reconstruction. Finally, a least squares approach was evaluated which found the best approximation to the overdetermined set of linear equations which define the desired sequence. Unfortunately, the results obtained from all three of these techniques were not encouraging. In particular, it was observed that signal to noise ratios in excess of 20db were required to obtain acceptable reconstructions. Furthermore, none on the algorithms performed significantly better than the others and, in addition, only slight improvements were made over the results obtained when no attempts were made to overcome the effects of noise.

The fundamental limitation with these methods is that any known or assumed statistics about the measurement noise is not factored into the reconstruction algorithm. As a first step towards an optimal reconstruction algorithm, the statistical characterization of signals corrupted by additive phase noise was developed. Specifically, it was assumed that the additive phase noise was a random process with known first and second order probability density functions. With this, the first and second order moments of the resulting signal were determined. In particular, these moments were found to be related to the characteristic functions of the probability density functions of the phase noise.

Publications Prepared During the First Year of the Grant

[1] P.L. Van Hove, M.H. Hayes, J.S. Lim, A.V. Oppenheim, "Signal Reconstruction From Fourier Transform Amplitude", accepted for publication in IEEE Transactions on Acoust., Speech, and Signal Processing.

[2] M.H. Hayes, "The Representation of Signals in Terms of Spectral Amplitude", to be presented at 1983 Int. Conf. on Acoust., Speech, and Sig. Proc., April 1983, Boston, MA.

[3] M.H. Hayes and T.F. Quatieri, "Recursive Phase Retrieval With Signal Boundary Conditions", submitted.

Graduate Student Activities

Although not specifically provided for in the budget, there has been considerable graduate student interest in the problem of reconstructing a signal from partial Fourier domain information. In particular, one Ph.D. student is presently actively involved in work pertaining to signal reconstruction from partial Fourier domain information in the presence of noise. In addition, a Master's student is presently investigating some exciting new ideas on extending the results of signal reconstruction from spectral amplitude.

II. SUMMARY OF WORK TO BE PERFORMED DURING THE UPCOMING YEAR

In the upcoming year, several issues outlined in [9] remain to be addressed. Specifically, in the second year of this grant, the work related to the developement of robust phase-only reconstruction algorithms will continue with an examination into the possibility of including noise statistics in the reconstruction algorithm. In addition, the incorporation of some other types of constraints in the iterative algorithms will be examined.

With respect to the phase retrieval problem, an analysis of the non-symmetric support constraint in the iterative phase retrieval algorithms [5] remains to be examined. In addition, the importance of the boundary values in the phase retrieval problem [3] will be examined in more detail. Also, regarding the amplitude and angle representation of signals, the convergence of the iterative algorithms remains an open question. Finally, the extension of the previous work to the more general problem of reconstructing signals given partial information about magnitude <u>and</u> phase will be initiated.

REFERENCES

[1-3] Shown on page 7.

[4] M.H. Hayes, J.S. Lim, and A.V. Oppenheim, "Signal Reconstruction From Phase or Magnitude", IEEE Trans. Acoust., Speech, and Signal Proc., vol. ASSP-28, pp. 672-680, Dec. 1980.

[5] M.H. Hayes, "The Reconstruction of a Multidimensional Sequence From the Phase or Magnitude of its Fourier Transform", IEEE Trans. Acoust., Speech, Sig. Proc., vol. ASSP-30, No. 2, pp. 140-154, April 1982.

[6] R.W. Gerchberg and W.O. Saxton, "A practical Algorithm for the Determination of Phase From Image and Diffraction Plane Pictures", Optik 35, pp. 237-246, 1972.

[7] A.V. Oppenheim, M.H. Hayes, J.S. Lim, "Iterative Procedures for Signal Reconstruction From Fourier Transform Phase", Optical Engineering, vol. 21, No. 1, pp. 122-127, Jan. 1982.

[8] C. Espy and J.S. Lim, "Effects of Noise on Signal Reconstruction From Fourier Transform Phase", Proc. 1982 Int. Conf. on Acoust., Speech, and Signal Proc., Paris, France, Vol. 3 pp. 1833-1836, May 1982.

[9] M.H. Hayes "Signal Reconstruction From Partial Phase and Magnitude Information", NSF Research Initiation Grant Proposal, School of Electrical Engineering, Georgia Institute of Technology, Nov. 1981.

CURRENT SUPPORT

[1] M.H. Hayes, "Signal Reconstruction From Partial Phase and Magnitude Information", Two year continuing Research Initiation Grant, NSF, Initiated 1 July, 1982, $48,000.

[2] M.H. Hayes, "Texture Segmentation and Boundary Detection in Images Using 2-D Linear Predictive Analysis", One year Joint Services Electronics Program Grant, USARO, Initiated April 1983, $29,530.

Washington, D.C. 20550

NSF FORM 98A

PLEASE READ INSTRUCTIONS ON REVERSE BEFORE COMPLETING

PART I—PROJECT IDENTIFICATION INFORMATION

1. Institution and Address	2. NSF Program	3. NSF Award Number
Georgia Institute of Technology School of Electrical Engineering Atlanta, GA. 30332	Research Initiation	ECS-8204793
	4. Award Period From 7/1/82 To 12/31/84	5. Cumulative Award Amount $48,000

6. Project Title

Signal reconstruction from Fourier transform phase or magnitude

PART II—SUMMARY OF COMPLETED PROJECT *(FOR PUBLIC USE)*

The goal of this research was to address some important questions and practical issues related to the reconstruction of a multidimensional signal from either the phase or the magnitude of its Fourier transform. Included in this work was an investigation into the importance of phase information in the representation of signals. This investigation led to a new way of representing spectral information, i.e., in terms of spectral *amplitude* and spectral *angle*. With such a representation it was shown that any causal finite length sequence is uniquely defined in terms of either its spectral amplitude or its angle. This result established a duality between amplitude and angle which resembles the duality found in the uniqueness of a minimum phase sequence in terms of its spectral magnitude and phase. Three different iterative algorithms were developed for reconstructing a signal from its spectral amplitude.

In order to investigate the robustness of the signal reconstruction problem in the presence of noise, several procedures for reconstructing a signal from noisy phase information were developed and studied. Specifically, noisy phase information was combined with knowledge of the true magnitude and/or hard constraints on the phase noise variations to define sets whose intersection contains the true Fourier transform values. Gerchberg/Saxton-type iterative algorithms for reconstruction from amplitude were implemented and studied.

The last part of this project was concerned with the importance of boundary conditions in the phase retrieval problem. In particular, the relationship between boundary conditions to off-axis holography was examined.

PART III—TECHNICAL INFORMATION *(FOR PROGRAM MANAGEMENT USES)*

1. ITEM *(Check appropriate blocks)*	NONE	ATTACHED	PREVIOUSLY FURNISHED	TO BE FURNISHED SEPARATELY TO PROGRAM: Check (√)	Approx. Date
a. Abstracts of Theses	XXXXX				
b. Publication Citations		XXXXX			
c. Data on Scientific Collaborators		XXXXX			
d. Information on Inventions	XXXXX				
e. Technical Description of Project and Results		XXXXX			
f. Other *(specify)*					

2. Principal Investigator/Project Director Name *(Typed)*	3. Principal Investigator/Project Director Signature	4. Date
Monson H. Hayes		7/24/85

NSF Form 98A (3-83) Supersedes All Previous Editions

Form Approved OMB No. 3145-0058

PART III: Technical Information

A. Publications

Journal Articles

[1] P.L. Van Hove, M.H. Hayes, J.S. Lim, and A.V. Oppenheim, "Signal Reconstruction From Signed Fourier Transform Magnitude," *IEEE Trans. on Acoustics, Speech, Sig. Proc.*, vol. ASSP-31, no. 5, pp. 1286-1293, Oct. 1983.

[2] M.H. Hayes and T.F. Quatieri, "Recursive Phase Retrieval Using Boundary Conditions," *Journal Opt. Soc. Am.*, vol. 73, no. 11, pp. 1427-1433, Nov. 1983.

Conference Papers

[1] M.H. Hayes, "The Representation of Signals in Terms of Spectral Amplitude," *Proc. 1983 Int. Conf. on Acoustics, Speech, and Sig. Proc*, pp. 1446-1449, April 1983.

[2] D.M. Thomas and M.H. Hayes, "Procedures for Signal Reconstruction From Noisy Phase," *Proc. 1984 Int. Conf. on Acoustics, Speech, and Sig. Proc*, pp. 31.1.1-31.1.4, March 1984.

B. Scientific Collaborators

1. David M. Thomas: Research Assistant (Graduate Student)

C. Technical Description of Project and Results

The goal of this research was to address some important questions and practical issues related to the reconstruction of a multidimensional signal from either the phase or the magnitude of its Fourier transform. In the next two sections a summary of the results obtained from this grant are presented. The first section discusses the results obtained on the reconstruction of a signal from its Fourier transform magnitude, which is the well-known *phase retrieval problem*. Following this are the results pertaining to the reconstruction of a signal from the phase of its Fourier transform which we refer to as the *magnitude retrieval problem*.

1. Phase Retrieval

The phase retrieval problem is concerned with the reconstruction of a signal given only knowledge of the magnitude of its Fourier transform. For discrete-time signals, the uniqueness of the solution to the phase retrieval problem has been recently considered in detail [1,2]. Specifically, it has been shown that for one-dimensional signals, spectral magnitude is not sufficient to uniquely define the signal. For multidimensional signals, on the other hand, the uniqueness is guaranteed if the signal has a multidimensional z-transform which is an irreducible polynomial (except for trivial ambiguities). In spite of this uniqueness, however, it is not generally possible to reconstruct a multidimensional signal from only its spectral magnitude. As part of this research, an investigation was undertaken to explain the observation that very little phase information seems to be required in order to reconstruct multidimensional signals from their spectral magnitude. This study was very fruitful and led to a better understanding of the theoretical issues pertaining to the phase and magnitude retrieval problems. Specifically, a new way of representing spectral information was established which involves writing the Fourier transform of a discrete time signal in terms of its *spectral amplitdue* and *spectral angle* rather than in terms of its spectral magnitude and phase [3]. Just as with magnitude and phase, Fourier transform amplitude and angle are both necessary in order to uniquely define a discrete-time signal. Spectral angle information is equivalent to the knowledge of the tangent of the phase whereas spectral amplitude is equivalent to knowledge of the magnitude of the Fourier transform along with one bit of phase information. With such a representation of spectral information, it was shown as a part of this research that any causal finite length sequence is uniquely defined in terms of its spectral amplitude [3,4]. Specifically,

Theorem: If $x(n)$ and $y(n)$ are two causal finite length sequences having the same spectral amplitude, then $x(n)=y(n)$.

With this result, a duality was established which resembles the duality found in the uniqueness of a minimum phase sequence in terms of its spectral magnitude and phase. Specifically,

Theorem: A sequence with a causal cepstrum is uniquely defined by the magnitude or to within a scale factor by the phase of its Fourier transform. Similarly, a causal (finite length) sequence is uniquely defined by the amplitude or to within a scale factor by the angle of its Fourier transform if *X(z)* contains no conjugate reciprocal zeros.

Once the uniqueness of a finite length signal in terms of its spectral amplitude and angle was established, the definition of spectral amplitude was then generalized to include a larger class of signal reconstruction problems. In particular, the one bit of phase information associated with the spectral magnitude was extended to include any division of the complex plane into two half planes [3].

Given a unique representation in terms of spectral amplitude and angle, the reconstruction of a signal from either amplitude or angle information was addressed. As has been previously demonstrated, reconstructing a signal from its spectral angle is straightforward [1,2]. The uniqueness of a signal in terms of its spectral amplitude, however, generally requires that the amplitude be known for all frequencies. Although only a finite number of amplitude samples are theoretically necessary, the specific frequencies for which the amplitude values are required cannot be determined *a priori*. Nevertheless, it has been experimentally observed that if the amplitude of the DFT of the signal is known and if the length of the DFT is large compared with the length of the signal then the solution is generally unique. Therefore, using an iterative approach similar to that introduced by Gerchberg and Saxton [5], three different iterative approaches were developed for reconstructing a signal from its amplitude. The difference between these algorithms lies in the method used to impose the known amplitude information in the iteration. In almost all cases, these algorithms sucessfully reconstructed the signal. However, a theoretical proof that the iteration will always converge to the correct signal (provided that the solution is uniquely defined by the given set of amplitude samples) has not yet been established.

Finally, an extension of some previous work on the importance of boundary conditions in the two-dimensional phase retrieval problem was undertaken [6]. Specifically, it was shown that specific geometries of point sources equivalently define the boundary conditions of a discrete two-dimensional field. These point source geometries represent a generalization of the conditions necessary for off-axis holography. Given the boundary conditions it is then generally possible to recover the two-dimensional signal from spectral magnitude information.

2. *Magnitude Retrieval*

The magnitude retrieval problem is concerned with the problem of reconstructing a signal from the phase of its Fourier transform. It has been recently shown that most finite length sequences are uniquely specified by the phase of their Fourier transform and a number of algorithms have been developed for reconstructing these finite length signals from their phase [1,2,7]. Unfortunately, however, it has been found that the reconstructed signals are very sensitive to noise. Therefore, due to its importance in applying phase-only reconstruction techniques to practical problems, part of the funded research has been directed towards an examination of the reconstruction errors which result from phase measurement errors.

In order to gain a better understanding of the issues involved, three *non-stochastic* algorithms were developed which attempt to make the signal reconstruction more robust in the presence of noise. These algorithms have been termed non-stochastic since they do not take into account any information or assumptions about the statistics of the phase noise. In all cases the underlying assumption was that noise corrupted measurements of the phase of a finite length sequence of length N were specified at kN distinct frequencies in the interval $(0,\pi)$. Since only N samples are required for phase-only signal reconstruction in the noise-free case, redundancy is provided by the excess number of amplitude samples. The first method which was investigated performs K reconstructions of $x(n)$ from N phase samples and then averages the reconstructed sequences [8]. The second method considers the possibility of smoothing or averaging the phase samples in order to obtain better phase estimates to use in the reconstruction. Finally, a least squares approach was evaluated which finds the best approximation to the overdetermined set of linear equations which define the desired sequence. The results from all three techniques were not encouraging. In particular, it was observed that signal to noise ratios in excess of 20db were required to obtain acceptable reconstructions. Furthermore, none of the algorithms performed significantly better than the others and, in addition, only slight improvements were made over the results obtained when no attempts were made to overcome the effects of noise.

In order to improve the reconstruction of a signal from noisy phase, additional constraints were considered [9]. In particular, measurements of noisy phase were combined with information about the true magnitude and/or hard constraints on the phase noise variations. Knowledge of spectral magnitude information along with noisy phase defines an arc in the frequency domain whereas a hard constraint on the phase noise variation defines a wedge in the frequency domain. Whatever the constraints, sets of legitimate frequency sample values are defined and a sequence with a Fourier transform consistent with this information was sought. Again, a Gerchberg-Saxton iterative procedure was used. The study concluded that the magnitude retrieval problem was very sensitive to phase measurement noise and that the more constraints one could apply in the iterative reconstruction process the more faithful the signal reconstruction.

REFERENCES

[1] M.H. Hayes, J.S. Lim, and A.V. Oppenheim, "Signal reconstruction from phase or magnitude", *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-28, pp. 672-680, Dec. 1980.

[2] M.H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform", *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-30, no. 2, pp. 140-154, April 1982.

[3] P.L. Van Hove, M.H. Hayes, J.S. Lim, and A.V. Oppenheim, "Signal Reconstruction From Signed Fourier Transform Magnitude," *IEEE Trans. on Acoustics, Speech, Sig. Proc.*, vol. ASSP-31, no. 5, pp. 1286-1293, Oct. 1983.

[4] M.H. Hayes, "The Representation of Signals in Terms of Spectral Amplitude," *Proc. 1983 Int. Conf. on Acoustics, Speech, and Sig. Proc*, pp. 1446-1449, April 1983.

[5] R.W. Gerchberg and W.O. Saxton, "A practical algorithm for the determination of phase from

image and diffraction plane pictures", *Optik 35*, pp. 237-246, 1972.

[6] M.H. Hayes and T.F. Quatieri, "Recursive Phase Retrieval Using Boundary Conditions," *Journal Opt. Soc. Am.*, vol. 73, no. 11, pp. 1427-1433, Nov. 1983.

[7] A.V. Oppenheim, M.H. Hayes, and J.S. Lim, "Iterative procedures for signal reconstruction from Fourier transform phase", *Optical Engineering*, vol. 21, no. 1, pp. 122-127, Jan. 1982.

[8] C. Espy and J.S. Lim, "Effects of additive noise on signal reconstruction from Fourier transform phase", *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-31, no. 4, pp. 894-898, Aug. 1983.

[9] D.M. Thomas and M.H. Hayes, "Procedures for Signal Reconstruction From Noisy Phase," *Proc. 1984 Int. Conf. on Acoustics, Speech, and Sig. Proc*, pp. 31.1.1-31.1.4, March 1984.

THE REPRESENTATION OF SIGNALS IN TERMS OF SPECTRAL AMPLITUDE*

M. H. Hayes

School of Electrical Engineering
Georgia Institute of Technology
Atlanta, Georgia 30332

ABSTRACT

In this paper, the importance of spectral phase and magnitude is examined from a different point of view. In particular, an amplitude and angle based representation of spectral information is developed. With this formulation, a causal finite length signal is uniquely defined by its spectral amplitude or, to within a scale factor, by its spectral angle.

INTRODUCTION

For both continuous-time and discrete-time signals, the magnitude and phase of the Fourier transform are, in general, independent functions, i.e., the signal cannot be recovered from knowledge of either one alone. With the appropriate a priori constraints, however, it is possible that either the spectral magnitude or the spectral phase may uniquely specify a signal. For example, when a signal is minimum phase or maximum phase, the log magnitude and phase are related through the Hilbert transform. For discrete-time sequences, it has also recently been shown that a finite-length sequence is uniquely specified to within a scale factor by its spectral phase assuming that the sequence contains no zero phase factors in the form of conjugate reciprocal zeros [1]. Unlike the minimum and maximum phase constraints, however, there is no dual statement of uniqueness between a sequence and its spectral magnitude under the same set of conditions. In particular, for any finite length sequence x(n) another finite length sequence with the same spectral magnitude may be easily created by the well-known procedure of "zero-flipping" [2].

In this paper, a different representation of spectral information is investigated. In particular, an amplitude and angle representation of Fourier transforms is developed. With such a representation, a causality constraint is sufficient for a discrete-time signal to be uniquely specified in terms of its spectral amplitude or, in most cases, to within a scale factor by its spectral angle. Although this uniqueness result may be easily extented to discrete samples of spectral angle, an arbitrary finite collection of spectral amplitude samples is not sufficient to uniquely define a causal finite length sequence. Nevertheless, sets of N spectral amplitude samples may be found which provide a unique characterization of a causal sequence of length N. Furthermore, if M is large enough, the spectral amplitude of the M-point DFT of a causal sequence of length N is sufficient for its unique characterization.

SPECTRAL AMPLITUDE AND ANGLE

Let x(n) denote a one-dimensional sequence and $X(\omega)$ its Fourier transform. For either real or complex-valued sequences, $X(\omega)$ is generally a complex-valued function of ω which may be written in polar form in terms of its magnitude and phase as:

$$X(\omega) = |X(\omega)|\exp[j\phi_x(\omega)] \tag{1}$$

where the phase, $\phi_x(\omega)$ is defined by

$$\phi_x(\omega) = \tan^{-1}[X_I(\omega)/X_R(\omega)] \tag{2}$$

and assumes values within the range $[-\pi, \pi]$. Note that, in addition to the value of the ratio $R(\omega)=X_I(\omega)/X_R(\omega)$, knowledge of $\phi_x(\omega)$ assumes that the sign of $X_R(\omega)$ and the sign of $X_I(\omega)$ are known for each frequency. Therefore, since

$$X_R(\omega) = |X(\omega)| \cos \phi_x(\omega) \tag{3a}$$

$$X_I(\omega) = |X(\omega)| \sin \phi_x(\omega) \tag{3b}$$

knowledge of $\phi_x(\omega)$ implies that the hard-clipped versions of $X_R(\omega)$ and $X_I(\omega)$ are known. It is the set of "zero-crossings" of $X_R(\omega)$ or of $X_I(\omega)$ which provide a key piece of information about $X(\omega)$ and, consequently, about x(n). For

* This work was a supported by the National Science Foundation under Grant ECS-8204793 and the Joint Services Electronics Program under contract DAAG29-81-K-0024.

example, it may be shown that $|X(\omega)|$ along with the "zero crossings" of $X_R(\omega)$ provide a unique specification of a finite duration causal sequence.

Instead of defining the phase of $X(\omega)$ as in (2), which presumes knowledge of the zero crossings of $X_R(\omega)$ and $X_I(\omega)$, suppose that the phase of $X(\omega)$ is defined by taking the principle value of the arctangent function in (2) so that it is confined to the range $[-\pi/2, \pi/2)$. Furthermore, let $\Phi_x(\omega)$ be used to denote this definition of the phase of $X(\omega)$ and let us refer to it as the angle of $X(\omega)$. Note that $\Phi_x(\omega)$ and $\tan[\phi_x(\omega)]$ are equivalent pieces of information about $X(\omega)$. Therefore, let $X(\omega)$ be written as

$$
\begin{aligned}
X(\omega) &= |X(\omega)| \exp[j\phi_x(\omega)] \\
&= |X(\omega)| \exp[j\Phi_x(\omega) + \psi_x(\omega)] \qquad (4) \\
&= A_x(\omega) \exp[\Phi_x(\omega)]
\end{aligned}
$$

where

$$A_x(\omega) = |X(\omega)| \exp[j\psi_x(\omega)] = |X(\omega)| \operatorname{sgn}[\cos \phi_x(\omega)] \qquad (5)$$

is defined to be the amplitude of $X(\omega)$. Note that $\psi_x(\omega)$ in (5) is equal to zero or π for each ω, i.e., $\tan[\psi_x(\omega)]=0$ for all ω. More specifically, $\psi_x(\omega)=0$ whenever $\phi_x(\omega)$ is within the interval $[-\pi/2, \pi/2)$ and it is equal to π otherwise. Therefore, the amplitude of $X(\omega)$ may equivalently be expressed as

$$A_x(\omega) = \begin{cases} |X(\omega)| & \text{if } -\pi/2 < \phi_x(\omega) < \pi/2 \\ -|X(\omega)| & \text{otherwise} \end{cases} \qquad (6)$$

Thus, spectral amplitude contains spectral magnitude information along with one bit of phase information, i.e., $\psi_x(\omega)$.

Finally it should be pointed out that whereas $|X(\omega)|$ is a continuous function of ω, $A_x(\omega)$ is discontinuous at those points where $\phi_x(\omega)$ passes through $\pm\pi/2$, i.e., at those frequences where $X_R(\omega)$ passes through zero. Thus, $A_x(\omega)$ contains information about both the magnitude of the transform as well as the zero crossings of the real part of the transform. For example, shown in Figure 1 is the magnitude, phase, amplitude, and angle of the Fourier transform of a discrete time signal of length N=4. Note that, as defined in (5), the amplitude of $X(\omega)$ is discontinuous at those frequencies where $\phi_x(\omega) = \pm\pi/2$ and is negative when the phase is outside the interval $[-\pi/2, \pi/2)$.

UNIQUENESS IN TERMS OF AMPLITUDE INFORMATION

Although a finite length sequence which has no zero phase component is uniquely defined to within a scale factor by its spectral phase, spectral magnitude does not place enough constraints on a finite length sequence to insure a unique solution. Specifically, if $x(n)$ is a finite length sequence with a z-transform $X(z)$ which has zeros at $z_1, z_2, \ldots, z_n$ then by replacing any one or more of these zeros with their conjugate reciprocals, i.e., replace the zero at $z=z_k$ with one at $z=1/z_k^*$ then the resulting sequence will have the same spectral magnitude. Although zero flipping necessarily preserves spectral magnitude, zero flipping must result in a sequence with a different spectral phase. A question of interest, therefore, is whether or not zero flipping results in a sequence with a different spectral amplitude. Without some additional information or constraints, however, this is not always the case. For example, consider an arbitrary finite length sequence $x(n)$ which has a spectral amplitude given by $A_x(\omega)$. With $y(n)=x(-n)$, note that the spectral magnitudes of $x(n)$ and $y(n)$ are the same. In addition, the spectral phases of $x(n)$ and $y(n)$ are related by:

$$\phi_x(\omega) = -\phi_y(\omega) \qquad (7)$$

Consequently, it follows that the zero crossings of the real parts of the transforms of $x(n)$ and $y(n)$ [the frequencies for which the spectral phase is equal to $\pm\pi/2$] are the same and, therefore, that the spectral amplitudes are identical. Causality, however, will eliminate this ambiguity and, in fact is a sufficient constraint for a finite length sequence to be uniquely defined by its spectral amplitude. More precisely:

> Theorem 1: If $x(n)$ and $y(n)$ are causal finite length sequences and if $A_x(\omega)=A_y(\omega)$ for all ω, then $x(n)=y(n)$.

Note that there is now a duality which is similar to the duality found in the uniqueness of a minimum phase sequence in terms of its spectral magnitude or spectral phase. Specifically, from Theorem 1 above and the uniqueness theorems concerning signal reconstruction from phase, the following corollary is now immediate:

> Corollary 1: A causal finite length sequence is uniquely defined by the amplitude of its Fourier transform and to within a scale factor by the angle of its Fourier transform if $X(z)$ contains no conjugate reciprocal zeros.

Note that although Theorem 1 is founded on a specific definition for the amplitude of the Fourier transform of a discrete-time signal, it is possible to adopt a more general definition. Specifically, note that the spectral amplitude is defined in (5) to be equal to its spectral magnitude when the phase is within the interval $[-\pi/2, \pi/2]$ and it is defined to be minus the spectral magnitude when the phase is outside this interval. As previously noted, with this definition of spectral amplitude, knowledge of the amplitude of the Fourier transform of a signal is equivalent to knowledge of the Fourier transform magnitude along with those frequencies for which the phase of the Fourier transform is equal to one of two possible values, $\pm \pi/2$. By choosing other values, different definitions for the amplitude may be obtained. For example, let ω_0 be an arbitrary number within the interval $[-\pi, \pi)$ and consider defining the amplitude of the Fourier transform of a discrete time signal as:

$$A_x(\omega:\omega_0) = \begin{cases} |X(\omega)| & \text{if } \omega_0<\phi_x(\omega)<\omega_0+\pi \\ -|X(\omega)| & \text{otherwise} \end{cases} \quad (8)$$

In this case, knowledge of $A_x(\omega:\omega_0)$ is equivalent to knowledge of the magnitude of $X(\omega)$ along with the frequencies for which the phase of $X(\omega)$ is equal to either ω_0 or $\omega_0+\pi$. Note that although Theorem 1 considers the case for which $\omega_0 = \frac{\pi}{2}$, it may be shown to hold for all values of ω_0 except for $\omega_0=0$. Specifically, [4]

Corollary 2: Let $x(n)$ and $y(n)$ be two causal finite length sequences. If $A_x(\omega:\omega_0)=A_y(\omega:\omega_0)$ for all ω with $\omega_0\neq 0$ then $x(n)=y(n)$.

UNIQUENESS IN TERMS OF AMPLITUDE SAMPLES

In the previous section, some uniqueness results were presented assuming that the spectral amplitude of a finite length sequence is known for all frequencies in the interval $[0,2\pi)$. In the case of spectral phase or spectral angle it is possible to generalize the uniqueness results to the case in which spectral phase or spectral angle is known only for a finite number of distinct frequencies. Specifically, it has been shown that for a finite length sequence of length N which has no symmetric (zero-phase) factors in its z-transform, any (N-1) samples of either its spectral phase or spectral angle is sufficient to uniquely define the sequence to within a scale factor [1]. Unfortunately, however, a finite set of amplitude samples is not always sufficient to uniquely specify a causal finite length sequence. For example, consider the following two causal sequences of length N=3

$$\begin{aligned} x(n) &= 1.0\delta(n) + 2.6\delta(n-1) + 1.2\delta(n-2) \\ y(n) &= 1.2\delta(n) + 2.6\delta(n-1) + 1.0\delta(n-2) \end{aligned} \quad (9)$$

Since $y(n)$ is obtained from $x(n)$ by flipping both of the zeros of $X(z)$ about the unit circle, both $x(n)$ and $y(n)$ have the same spectral magnitude. Furthermore, in the interval $(0, \pi)$, the real part of the Fourier transform of $x(n)$ is equal to zero at only one frequency, $\omega = .477023\,\pi$ and the real part of the Fourier transform of $y(n)$ is equal to zero only at $\omega=.526166\,\pi$. Therefore, the amplitude of $X(\omega)$ is equal to the amplitude of $Y(\omega)$ for all ω outside the intervals $(.477023\pi, .526166\pi)$ and $(-.526166\pi,-.477023\pi)$. Consequently, an arbitrary number of amplitude samples within this region is not sufficient to distinguish $x(n)$ from $y(n)$. Note, however, that one sample of the amplitude within the interval $(.477023\,\pi, .526166\,\pi)$ is sufficient to distinguish $x(n)$ from $y(n)$.

Thus, although a given set of samples will not lead to a unique specification of a sequence in terms of spectral amplitude in all cases, it may be shown that a finite set of samples may always be found which provide this unique specification [4]. In particular, any causal finite length sequence of length N may be uniquely defind by the spectral magnitude of its M-point DFT provided M is chosen large enough.

RECONSTRUCTION FROM AMPLITUDE

In this section, the problem of reconstructing a finite length sequence from the amplitude of its Fourier transform is addressed. As previously discussed, a given finite set of amplitude samples is not always sufficient to uniquely specify the sequence. In this section, however, it will be assumed that the unknown sequence, $x(n)$, is zero outside the interval $[0,N-1]$ and that the amplitude of its M-point DFT, $A_x(k)$, is known and that M is large enough to insure a unique specification of $x(n)$.

Motivated by the iterative algorithm originally proposed by Gerchberg and Saxton [5], an iterative procedure has been used in the reconstruction from amplitude problem. Specifically, the problem may be viewed as one in which some signal constraints are known both in the time and frequency domains. In particular, in the time domain $x(n)$ is known to have its support confined to the interval $[0,N-1]$ and in the frequency domain it is known to have an M-point DFT with amplitude $A_x(k)$. The iteration is thus characterized by the repeated transformation between the time and frequency domains where in each domain and at each step in the iteration the signal constraints are imposed on the current estimate. The incorporation of the time domain constraint is straight-forward since it involves simply a windowing operation. There are several alternatives, however, for imposing the frequency domain constraint. Specifically, with $x_\ell(n)$ the estimate obtained after ℓ iterations, let $X_\ell(k)$ be its M-point DFT which has an amplitude given by $A_\ell(k)$. The frequency domain constraint to be placed on $X_\ell(k)$ is the known spectral amplitude of $x(n)$, $A_x(k)$. In the complex plane, the DFT of $x_\ell(n)$ is thus constrained to lie on the semicircle defined by the intersection of a circle of radius $A_x(k)$ and the half plane of all positive real parts if $A_x(k)>0$ or the half plane of all negative real parts if $A_x(k)<0$. Thus, the known amplitude imposes both a magnitude as well as a phase constraint on $X_\ell(k)$. The flexibility in incorporating the amplitude information lies in the method by which the phase information is imposed. Since there is no reason for altering the phase of $X_\ell(k)$ if it lies within the correct interval, the only question is what phase should be used for $X_\ell(k)$ when the phase of $X_\ell(k)$ falls outside the given interval. One possibility is to set the phase to zero if the phase of $X(k)$ is known to lie in the interval $[-\pi/2, \pi/2]$ and to set it equal to π otherwise.

Another possibility for incorporating the given amplitude information is to set the amplitude of $X_{\ell+1}(k)$ equal to the known amplitude, $A_x(k)$:

$$A_{\ell+1}(k) = A_x(k) \qquad (10)$$

With this approach, $X_\ell(k)$ is scaled so that it has the correct magnitude and then, if necessary, a phase of π is added to $X_\ell(k)$. However, if the real part of $X(k)$ is close to zero and the sign of the real part of $X_\ell(k)$ differs from that of $X(k)$, then the incorporation of the amplitude constraint (10) will significantly increase the error between $X(k)$ and $X_{\ell+1}(k)$. Another possibility, therefore, is to simply scale $X_\ell(k)$ so that it has the correct magnitude and then set the sign of the real part of $X_{\ell+1}(k)$ equal to the sign of the real part of $X_\ell(k)$.

With either of these last two approaches for imposing the frequency domain constraint, the iterative procedure has been observed to converge, in most cases, to the correct sequence when $x(n)$ is uniquely defined by the amplitude of its M-point DFT. A theoretical proof of convergence, however, has not yet been obtained. Although the number of iterations required to reach a convergent solution is in general very large, this number tends to decrease as the length of the DFT is increased.

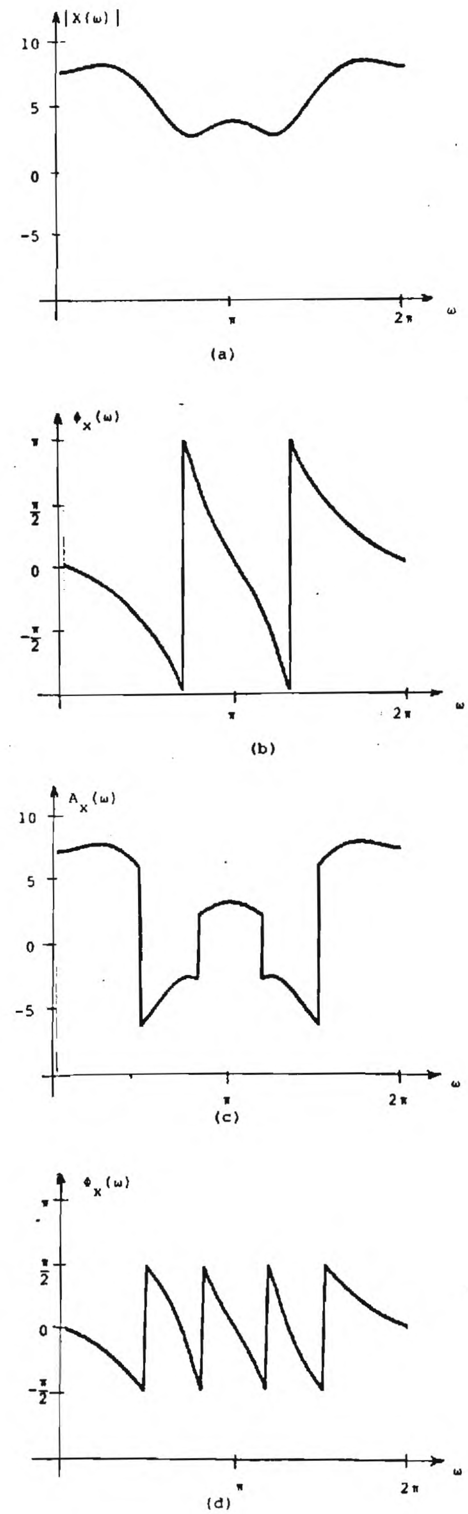


Figure 1: (a) Spectral magnitude, (b) Spectral phase (c) Spectral amplitude, (d) Spectral angle of a sequence of length $N = 4$.

REFERENCES

[1] M.H. Hayes, J.S. Lim, and A.V. Oppenheim, "Signal Reconstruction From Phase or Magnitude," IEEE Trans. Acoust., Speech, and Signal Processing, vol. ASSP-28, pp.672-680, Dec. 1980.

[2] L.S. Taylor, "The Phase Retrieval Problem," IEEE Trans. Ant. and Prop., vol. AP-29, no. 2, pp. 386-391, March 1981.

[3] M.H. Hayes, "The Reconstruction of a Multidimensional Sequence From the Phase or Magnitude of its Fourier Transform," IEEE Trans. Acoust., Speech, and Signal Proc., Vol. ASSP-30, No. 2, April 1982, pp. 140-154.

[4] P. L. Van Hove, M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal Reconstruction from Fourier Transform Amplitude," submitted to IEEE Trans. on Acoust. Speech, and Sig. Proc.

[5] G.W. Gerchberg and W.O. Saxton, "A Practical Algorithm for the Determination of Phase From Image and Diffraction Plane Pictures," Optik 35, pp.237-246, 1972.

PROCEDURES FOR SIGNAL RECONSTRUCTION FROM NOISY PHASE*

D. M. Thomas and M. H. Hayes

Georgia Institute of Technology
School of Electrical Engineering
Atlanta, Georgia 30332

ABSTRACT

In this paper a class of iterative procedures is presented for reconstructing a finite duration sequence from noisy samples of its Fourier transform phase. These measurements are combined with a knowledge of the true transform magnitude and/or hard constraints on the phase noise variations to define sets whose intersection must contain the true sequence. The algorithms iterate between the sequence domain and the transform domain applying the known constraints (i.e. finite duration and known limits on phase variation) in each domain. Results of an experimental investigation are presented. A method is described for the case where limits on both the magnitude and phase variation of a finite length sequence are known.

INTRODUCTION

In recent years the problem of signal restoration from partial or incomplete information has received considerable attention. In the magnitude retrieval problem, it is assumed that the Fourier transform phase, $\phi(\omega) = \arg[X(\omega)]$, of a finite duration sequence is known exactly at N-1 distinct frequencies in the interval $(0,\pi)$ where N is the known extent of the sequence. Under these conditions and with certain restrictions on the placement of the zeros of X(z), it has been shown [1] that iterative procedures exist which will converge to the unique solution. These algorithms iterate between the sequence domain and the transform domain applying the known constraints (i.e. finite duration and known phase) in each domain. Convergence has been proven within the framework of nonexpansive mapping theory [2].

In contrast, recovering a finite duration sequence x(n) from a knowledge of its Fourier transform magnitude alone has proven to be a much more formidable problem and few sequences meet the uniqueness criterion presently available [3]. This situation has prompted a number of investigators to study the reconstruction of a sequence from samples of its signed magnitude (i.e. the transform magnitude and one bit of phase information) [4]. In this study it has been shown that if x(n) and y(n) are two real, causal (or anticausal), finite duration sequences and if certain restrictions on the placement of the zeros of X(z) are met, then x(n) is equal to y(n) if their respective signed magnitudes are equal at all frequencies. Unfortunately, a given finite set of samples of the signed magnitude is not always sufficient to uniquely specify a sequence. This is true even through it is known that a maximum number of 3N suitably chosen frequency samples suffice to uniquely specify a sequence. The key words are "suitably chosen" since the required samples vary from sequence to sequence. In practice, picking a sufficiently long transform length (typically 10 times the length of the sequence) allows excellent restorations.

When the phase of $X(\omega)$ is not known accurately or when it is corrupted by noise, it has been observed that the magnitude retrieval procedure described above may perform badly. Other techniques, such as reconstruction averaging [5] or minimum cross-entropy methods [6] may also perform poorly, particularly in low signal to noise ratio environments. It seems likely then, that additional information is required if we are to achieve more acceptable performance.

In this paper, a new class of algorithms is presented which achieve significantly better reconstructions over those obtained in the past. These procedures utilize a knowledge of the true transform magnitude and/or bounds on the phase noise variation to define constraint sets whose intersection contains the undistorted signal. At each stage of the iteration, the current estimate of x(n) is projected (in some fashion) onto the constraint sets. In all cases to be examined, the finite duration requirement is enforced in the sequence domain.

Signal Restoration - Noise Free Phase

Let x(n) denote a finite duration one-dimensional sequence which is zero outside the the interval $0 \leq n \leq N-1$. The z-transform and the Fourier transform of x(n) will be denoted by X(z)

*This work was supported by the National Science Foundation under Grant ECS-8204793 and by the Joint Services Electronics Program under contract DAAG29-81-K-0024.

and $X(\omega)$. Since $X(\omega)$ is, in general, a complex-valued function of ω, it may be written in terms of its real and imaginary parts, $X_R(\omega)$ and $X_I(\omega)$ respectively, or in terms of its magnitude and phase, $|X(\omega)|$ and $\phi_x(\omega)$ respectively, as follows:

$$X(\omega) = X_R(\omega) + jX_I(\omega) = |X(\omega)|e^{j\phi_x(\omega)}$$

with: $\quad \phi_x(\omega) = \tan^{-1}\left[X_I(\omega)/X_R(\omega)\right].$

A set of conditions under which $x(n)$ may be recovered from samples of its Fourier transform phase is contained in the following theorem [1]:

> <u>Theorem</u>: Let $x(n)$ be a sequence which is zero outside the interval $0 \le n \le N-1$ with $x(0) \ne 0$ and which has a z-transform with no zeros on the unit circle or in conjugate reciprocal pairs. Then samples of the Fourier transform phase $\phi_x(\omega)$ at $N-1$ distinct frequencies in the interval $0<\omega<\pi$ suffice to uniquely (i.e. to within a positive constant) define $x(n)$.

A special case of this theorem arises when the phase samples of $X(\omega)$ are equally spaced around the unit circle such as occurs in the Discrete Fourier Transform (DFT). Let M be the length of the DFT. Then, if $M>2N$, the iterative procedure, which replaces the estimates in each domain by their known values, will converge to the unique solution $x(n)$ for any initial starting point $x_o \in R^N$ [2]. It is this algorithm that motivates our methods for dealing with noisy phase measurements.

Signal Restoration - Noisy Phase

When noise is added to the phase measurements, the phase-only iterative procedure described above generally produces poor reconstructions. Approaches which use restoration averaging [5] or minimum cross entropy methods [6] also seem to give poor results. This is especially true of these methods in low signal to noise ratio cases. In this section, we will present a number of iterative schemes useful in reconstructing a sequence given noisy phase measurements. These methods have been found to give significantly better reconstructions than have been previously obtained albeit at the cost of higher information requirements. Here the added cost takes the form of knowledge of the true magnitude and/or hard constraints on the phase noise variation. As in the noise-free case, the iteration proceeds by projecting the current estimate of $x(n)$ onto the constraint sets in both the sequence domain and the transform domain. In all cases to be examined, the constraint set in the sequence domain is the knowledge of the first (non-zero) point and the finite duration requirement. The four methods to be described differ only in what constitutes the constraint set in the transform domain.

Given that $\omega_k = 2\pi k/M$, $k = 0,1, \ldots ,M-1$ are a set of M distinct frequencies with $M>2N$, each of the projection operators can be described as follows (see Figure 1 for illustrations of the constraint sets defined below.):

- <u>Method I</u>: No magnitude information is available and the phase noise is known to vary no more then $\pm \Delta_p$ radians from its true value. Alternatively, we can say that the true phase is no more than $\pm \Delta_p$ radians distant from the measured noisy value. At each stage of the iteration, we examine the phase estimate $\phi(\omega_k)$. If it is more than Δ_p radians distant from the measurement value we replace the estimate with the measurement. The magnitude is not altered.

- <u>Method II</u>: Here, the true transform magnitude $|X(\omega_k)|$ is known for all values of k, but, no information is available on the variability of $\phi(\omega_k)$. We do, however, have the noisy phase measurements. For each value of k and at each stage of the iteration we replace the magnitude estimate by its true value. The phase estimate, however, is not changed. It should be pointed out that this procedure differs from the phase retrieval problem, as it usually takes form, in that an initial, if noisy, estimate of the phase is available.

- <u>Method III</u>: Knowledge of the true magnitude is combined in this method with a hard constraint Δ_p on the phase noise variation to define an arc centered on the measured phase value. At each stage of the iteration, the current estimate of $X(\omega_k)$ is projected, using a nearest neighbor rule, onto the convex hull of these arcs.

- <u>Method IV</u>: As in the third method, the true magnitude $|X(\omega_k)|$ is known as is the maximum phase noise variation Δ_p. However, at each stage of the process, the magnitude estimate is replaced by the true value of the magnitude. The phase estimate is replaced by the noisy phase measurement if the estimate lies more than Δ_p radians away from the phase measurement.

To test these four methods, an eight-point non minimum phase sequence $\{x(n) = (4,2,-11,5,4,5,15,-6)\}$ was used. An M-point DFT of this sequence is calculated and to it's phase is added an M-point noise sequence. This noise sequence is generated using a uniform probability density function given by:

$$p_{\Delta_p}(w) = \begin{cases} \frac{1}{2\Delta p}, & -\Delta_p < w < \Delta_p, \\ 0, & \text{otherwise.} \end{cases}$$

Each of the four methods were examined for transform lengths of $M = 16$, 32, 64 and 128 points and noise variations Δ_p equal to $\pi/2$, $10^{-1}\pi$, $10^{-2}\pi$, $10^{-3}\pi$, $10^{-4}\pi$, and $10^{-5}\pi$. A relaxation parameter [2] of 0.99 is used in these experiments. Two performance measures were calculated. They are 1) the normalized mean square error (NMSE) defined by [5]

$$\text{NMSE} = \frac{\sum_{n=0}^{N-1} [x(n) - Ax_f(n)]^2}{\sum_{n=0}^{N-1} x^2(n)},$$

and 2) the mean square error improvement ratio (MSEI):

$$\text{MSEI} = \frac{\sum_{n=0}^{M-1} [x(n) - x_o(n)]^2}{\sum_{n=0}^{M-1} [x(n) - Ax_f(n)]^2}.$$

In these expressions, $x(n)$ is the true sequence, $x_o(n)$ is the initial or starting estimate of $x(n)$, $x_f(n)$ is the final estimate of $x(n)$, and A is a parameter chosen to minimize the normalized mean square error.

A number of observations can be made from the results obtained thus far in our investigations:

- Methods II and IV perform the best in the noise ranges $10^{-1}\pi$ to $10^{-5}\pi$. For noise variations greater than $10^{-1}\pi$, Method IV is the algorithm of choice. For example, if $\Delta_p = \pi/2$, Method II obtains only 21 dB. of mean square error improvement. Method IV, on the other hand, gives a nearly perfect reconstruction with an MSEI of 242 dB. for M equal to 64 points and 5000 iterations

- At intermediate noise levels (i.e. $10^{-1}\pi$ to $10^{-5}\pi$) methods II and IV give roughly equivalent results. However, as M becomes larger, Method IV tends to converge much more quickly. In the same range of noise levels, Method II and IV converge at the same rate as the phase-only iteration using the sequence without noise added.

- Method III obtains large reductions in the error in the early stages but very little improvement (if any) is obtained as the iteration proceeds.

- Method I is the least effective of the four methods examined. However, it also requires the least amount of a priori information about $x(n)$ or its transform.

- As a rule, increasing M, the length of the transform, improves the convergence rate of the algorithms. However, in some instances, a shorter DFT length can outperform a longer length.

Figure 2 illustrates these observations concerning the four algorithms. Here, M is equal to 64, and for those sequences with noise present it is given a maximum variation of $10^{-3}\pi$ radians. Also shown in this figure is the performance of the phase-only iteration for the phase samples with no phase noise added and with a noise level of $\Delta_p = 10^{-3}\pi$ added. The differences between the initial mean square errors are attributable to whether or not magnitude information is available.

Shown in Figure 3, is the performance of Method IV as Δ_p is varied. Similar effects are seen in the results from the other procedures. In this figure, M is equal to 64.

Signal Restoration - Noisy Phase and Magnitude

In this section, we relax the need for knowledge of the true magnitude and require instead that bounds on the variations in phase and magnitude noise be known. These bounds define a region about the measured magnitude and phase which must contain the true transform values of the sequence. At each stage of the iteration we project, using a nearest neighbor rule, the estimate of $X(\omega_k)$ onto the convex hull of the region described above. Again, in the sequence domain we enforce the finite duration requirement. Also, we assume $x(0)$ is known.

Figure 4 represents the results of an experiment using the eight-point sequence defined earlier, Δ_p equal to $10^{-3}\pi$, and M equal to 64. Various levels of uniformly distributed noise ranging from 10^{-1} down to 10^{-5} were added to the magnitude data and reconstructions were obtained as shown. Also shown is the behavior of the phase-only iteration when given the true sequence's phase and when given the true phase corrupted by phase noise at a level of $10^{-3}\pi$.

REFERENCES

[1] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal Reconstruction from Phase or Magnitude," IEEE Trans. on Acoust., Speech, and Sig. Proc., Vol. ASSP-28, No. 6, Decmeber 1980, pp. 672-680.

[2] V. T. Tom, T. F. Quatieri, M. H. Hayes, and J. H. McClellan, "Convergence of Iterative Non-Expansive Signal Reconstruction Algorithms," IEEE Trans. on Acoust., Speech, and Sig. Proc., Vol. ASSP-29, No. 5, October 1981, pp. 1052-1058.

[3] A. V. Oppenheim, J. S. Lim, and S. R. Curtis, "Signal Synthesis and Reconstruction from Partial Fourier-Domain Information," J. Opt. Soc. Am., Vol. 73, No. 11, November 1983, pp. 1413-1420.

[4] P. L. Van Hove, M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal Reconstruction From Signed Fourier Transform Magnitude," IEEE Trans. on Acoust., Speech, and Sig. Proc., Vol. ASSP-31, No. 5, October 1983, pp. 1286-1293.

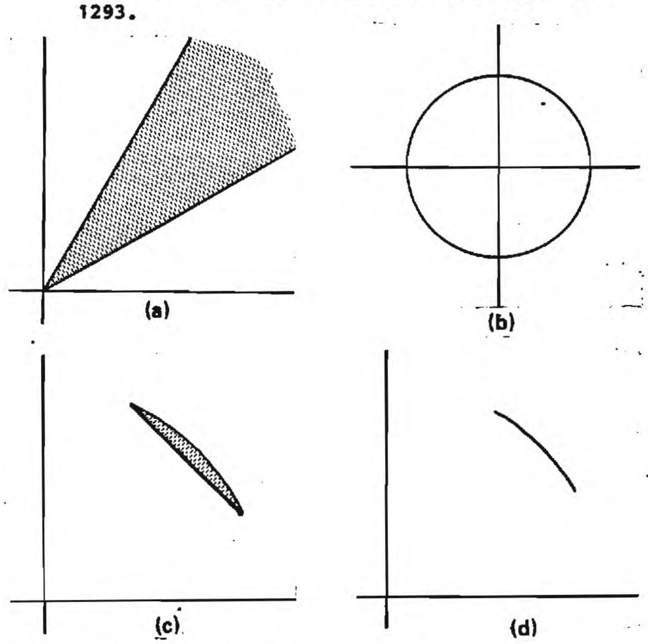


FIGURE 1. CONSTRAINT SETS FOR THE FOUR METHODS DISCUSSED IN THE PAPER: (a) method I – true sequence is known to lie in the intersection of M wedge-shaped regions, (b) method II – true sequence lies in the intersection of M circles, (c) method III – true sequence lies in the intersection of the convex hulls of M arcs, and (d) method IV – true sequence is known to lie in the intersection of M arcs.

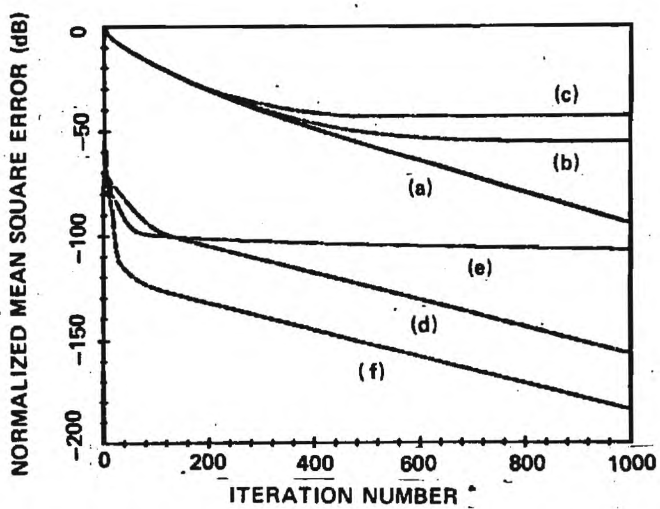


FIGURE 2. COMPARISON OF METHODS (N=8,M=64): (a) phase only iteration – no noise, (b) phase only iteration – noise level of 0.001π, (c) method I – noise level of 0.001π, (d) method II – noise level of 0.001π, (e) method III – noise level of 0.001π, (f) method IV – noise level of 0.001π.

[5] C. Y. Espy and J. S. Lim, "Effects of Additive Noise on Signal Reconstruction from Fourier Transform Phase," IEEE Trans. on Acoust., Speech, and Sig. Proc., Vol. ASSP-31, No. 4, August, 1983, pp. 894-898.

[6] B. Musicus, "Iterative Algorithms for Optimal Signal Reconstruction and Parameter Identification Given Noisy and Incomplete Data," Ph.D. Thesis, M.I.T., Dept. of EECS (Aug. 1982).

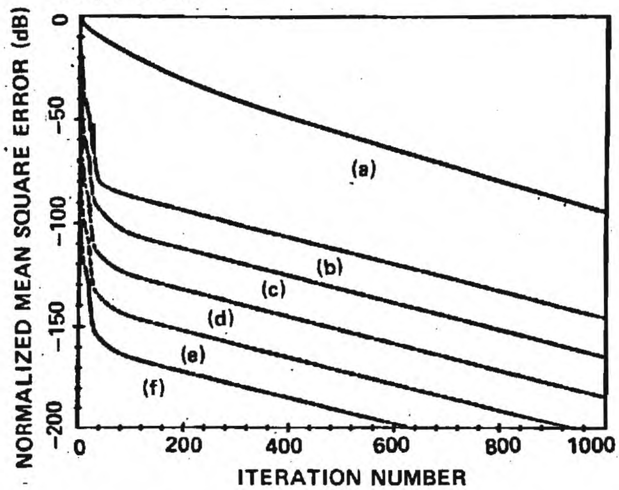


FIGURE 3. EFFECTS OF NOISE LEVEL ON PERFORMANCE USING METHOD IV (N=8,M=64): (a) phase only iteration – no noise, (b) noise level of 0.1π, (c) noise level of 0.01π, (d) noise level of 0.001π, (e) noise level of 0.0001π, (f) noise level of 0.00001π.

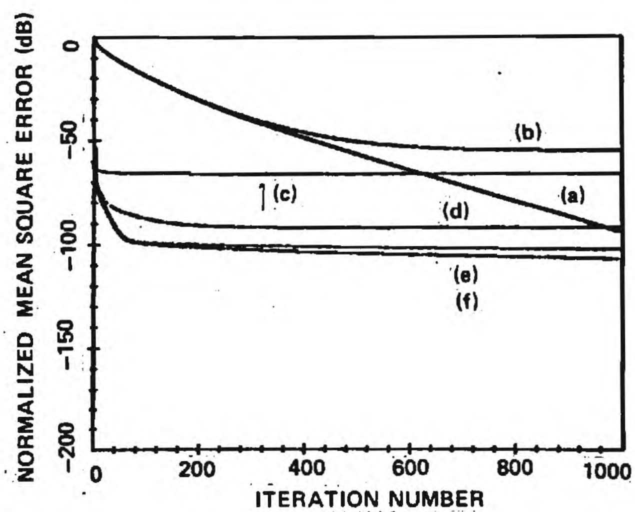


FIGURE 4. RECONSTRUCTION FROM NOISY PHASE AND MAGNITUDE USING CONVEX HULL PROJECTIONS: (a) phase only procedure – no noise, (b) phase only iteration – phase noise level of 0.001π, (c) magnitude noise level of 0.1, (d) magnitude noise level of 0.001, (e) magnitude noise level of 0.00001, and (f) no magnitude noise. (note: unless otherwise specified there is a phase noise level of 0.001π.)

Reprinted from Journal of the Optical Society of America, Vol. 73, page 1427, November 1983

Recursive phase retrieval using boundary conditions

Monson H. Hayes

School of Electrical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332

Thomas F. Quatieri

Massachusetts Institute of Technology, Lincoln Laboratory, Lexington, Massachusetts 02173

Received March 31, 1983; revised manuscript received June 16, 1983

The phase-retrieval problem for discrete multidimensional fields is investigated. In particular, a recursive procedure is developed for reconstructing a signal from the modulus of its Fourier transform. The information necessary to begin the recursion is the boundary values of the signal. Although it is not always possible to determine these boundary values from Fourier modulus data only, if the sequence has a region of support with a certain geometry then these boundary values can be determined. These geometries represent a generalization of the conditions for off-axis holography.

1. INTRODUCTION

The reconstruction of a signal from the magnitude of its Fourier transform, generally referred to as the phase-retrieval problem, arises in a variety of different contexts and applications and within such diverse fields as crystallography, astronomy, optics, and signal processing.[1,2] There are three fundamental issues involved in the phase-retrieval problem: the uniqueness of the solution, the development of algorithms for reconstructing a signal from the magnitude of its Fourier transform, and the sensitivity of the reconstruction to measurement errors and computational noise. In this paper attention is focused on the reconstruction problem. More specifically, following a brief review in Section 2 of some recent results concerning the uniqueness of the solution to the phase-retrieval problem for discrete two-dimensional signals, a recursive solution to the phase-retrieval problem is developed in Section 3. This recursive algorithm is similar to other phase-retrieval algorithms in the sense that some signal information, other than the magnitude of its Fourier transform, is assumed to be known.[1–9] Specifically, this recursive algorithm assumes knowledge of what we presently define as the boundary values of the signal. Although it is not always the case that the boundary values of a two-dimensional signal are known, it is shown in Section 4 that, in some cases, the boundary values of a signal may be determined from the given Fourier-transform magnitude information. In particular, it is shown that if a two-dimensional sequence has a region of support with a certain geometry, then the boundary values of the sequence may be easily recovered. These geometries represent a generalization of the conditions for off-axis holography.

2. PHASE RETRIEVAL

In order to develop the recursive phase-retrieval algorithm in Section 3, some notation and terminology related to discrete two-dimensional signals are necessary. The required background is therefore provided in Section 2.A. In addition, some recent results concerning the uniqueness of the solution to the phase-retrieval problem are briefly reviewed in Section 2.B.

A. Notation and Terminology

A two-dimensional sequence is a function of two integer variables m and n, which is denoted by $x(m, n)$. The two-dimensional z transform of $x(m, n)$ is denoted by $X(z_1, z_2)$ and is defined by

$$X(z_1, z_2) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(m, n) z_1^{-m} z_2^{-n}, \tag{1}$$

where z_1 and z_2 are complex variables. The two-dimensional Fourier transform of $x(m, n)$ is equal to the z transform of $x(m, n)$ evaluated along the unit bi-disk $|z_1| = |z_2| = 1$ and is given by

$$X(e^{j\omega_1}, e^{j\omega_2}) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} x(m, n) e^{-j\omega_1 m} e^{-j\omega_2 n}, \tag{2}$$

where ω_1 and ω_2 are real variables that represent the spatial frequencies of the two-dimensional Fourier transform. Written in polar form, $X(e^{j\omega_1}, e^{j\omega_2})$ is expressed in terms of its magnitude and phase as

$$X(e^{j\omega_1}, e^{j\omega_2}) = |X(e^{j\omega_1}, e^{j\omega_2})| e^{j\phi_x(\omega_1, \omega_2)}. \tag{3}$$

Thus the phase-retrieval problem is concerned with the recovery of $x(m, n)$ given only the spectral magnitude function $|X(e^{j\omega_1}, e^{j\omega_2})|$.

The two-dimensional sequences considered in this paper are assumed to be real valued and to have finite support, i.e., $x(m, n)$ is real and nonzero for only a finite number of values of the ordered pair (m, n). For convenience it is assumed, without any loss in generality, that a sequence with finite support has first quadrant support, i.e., $x(m, n) = 0$ if $m < 0$ or if $n < 0$. In addition, if it is known that $x(m, n)$ is zero outside the rectangular region $R(M, N)$ containing all points (m, n) for which $0 \le m < M$ and $0 \le n < N$, i.e.,

$$R(M, N) = [0, M-1] \times [0, N-1], \quad (4)$$

where $\times$ is used to denote the Cartesian cross products, e.g., then $x(m, n)$ is said to have support $R(M, N)$.

Since the two-dimensional z transform of a sequence with first quadrant support is a polynomial in the two variables, z_1^{-1} and z_2^{-1}, $X(z_1, z_2)$ may always be uniquely written (to within factors of zero degree) as a product of polynomials that are irreducible over the field of complex numbers[10]

$$X(z_1, z_2) = \alpha z_1^{-n_1} z_2^{-n_2} \prod_{k=1}^{p} X_k(z_1, z_2), \quad (5)$$

where α is a real number and n_1 and n_2 are nonnegative integers. The irreducible factors $X_k(z_1, z_2)$, which may be of arbitrarily large degree, are the two-dimensional counterpart of the linear factors that define the zeros of the z transform of a one-dimensional sequence. Clearly, since these irreducible factors are polynomials in two variables, the zero sets of two-dimensional z transforms are contours in the z_1–z_2 plane.

B. Uniqueness

An important issue in the phase-retrieval problem is the uniqueness of the solution. It is well known that, without any additional information or constraints, a signal (discrete or continuous, one-dimensional or multidimensional) is not uniquely specified by the magnitude of its Fourier transform.[1–4,11] The absence of a unique solution stems from the fact that it is always possible to convolve a signal with an arbitrary all-pass signal (one that has a Fourier transform with unit modulus) to obtain another signal with the same spectral magnitude. As a result, the ability to incorporate some additional information or knowledge about the signal to constrain the set of admissible solutions is necessary in order to obtain a unique reconstruction. Since many of the signals that are of practical interest are of finite duration or extent, a finite support constraint is often used in phase-retrieval algorithms.[1–4] As a result, the uniqueness of the solution to the phase-retrieval problem has been considered for the case in which the solution is constrained to be of finite length or to have finite support. Unfortunately, however, it has been shown that for one-dimensional signals (either continuous or discrete) such a constraint is not sufficient to ensure a unique solution because of the possibility of zero flipping.[2,3,11] For two-dimensional signals with finite support, on the other hand, the uniqueness results are considerably different. Although the uniqueness properties are not well understood for the continuous case, considerable progress has been made for the discrete case. In particular, it has been shown that the two-dimensional counterpart of zero flipping in the discrete one-dimensional case is the flipping of the zero contours of the irreducible polynomials that define the two-dimensional z transform of the sequence.[3] It follows therefore that, if $X(z_1, z_2)$ is an irreducible polynomial, then $x(m, n)$ is uniquely defined by its spectral magnitude to within the trivial ambiguities of a linear shift, a reflection of the sequence about the origin, or by a scale factor of (-1). More specifically, note that, if two sequences $x(m, n)$ and $y(m, n)$ are related by

$$y(m, n) = \pm x(\pm m + k, \pm n + l) \quad (6)$$

for some integers k and l, then $x(m, n)$ and $y(m, n)$ have Fourier transforms with the same magnitude. Therefore any two sequences related by Eq. (6) are said to be equivalent, and this equivalence relation is denoted by

$$x(m, n) \sim y(m, n). \quad (7)$$

With this relation, the uniqueness result of interest is the following[3]:

Theorem 1: Let $x(m, n)$ be a two-dimensional sequence with finite support that has a two-dimensional z transform that, except for trivial factors of the form $az_1^{-k_1}z_2^{-k_2}$, is irreducible. If $y(m, n)$ is another two-dimensional sequence with finite support with $|Y(e^{j\omega_1}, e^{j\omega_2})| = |X(e^{j\omega_1}, e^{j\omega_2})|$ for all ω_1 and ω_2, then $y(m, n) \sim x(m, n)$.

It should be pointed out that the requirement that $X(z_1, z_2)$ be irreducible is not a particularly strong constraint. Specifically, it may be shown that within the set of all two-dimensional sequences with finite support, the subset of all sequences that have reducible z transforms is a set of measure zero.[3,12] As a result, almost all two-dimensional sequences with finite support will satisfy the irreducibility requirement of Theorem 1. Irreducibility of the z transform of a two-dimensional sequence may, in fact, be guaranteed with the proper placement of point sources outside the sequence's region of support.[13]

One limitation of Theorem 1 is that it requires that the magnitude of the Fourier transforms of $x(m, n)$ and $y(m, n)$ be equal for all values of ω_1 and ω_2. Fortunately, however, Theorem 1 may be extended so that the magnitudes of the Fourier transforms of $x(m, n)$ and $y(m, n)$ need only be equal for a finite number of values of ω_1 and ω_2. The number of points for which the Fourier-transform magnitudes must be equal is determined by the size of the regions of support of $x(m, n)$ and $y(m, n)$, whereas the locations of the sample points in the ω_1–ω_2 plane are constrained to lie on a regular lattice. Specifically[3]:

Theorem 2: Let $x(m, n)$ and $y(m, n)$ be two-dimensional sequences with support $R(M, N)$. If a_k for $k = 1, \ldots, M$ and b_l for $l = 1, \ldots, N$ are distinct real numbers in the interval $(0, \pi)$ and if

$$|X(e^{j\omega_1}, e^{j\omega_2})| = |Y(e^{j\omega_1}, e^{j\omega_2})| \quad \text{for} \quad \begin{array}{l} \omega_1 = a_1, a_2, \ldots, a_M, \\ \omega_2 = b_1, b_2, \ldots, b_N \end{array} \quad (8)$$

then $y(m, n) \sim x(m, n)$.

A special case of this theorem results when the points a_k and b_l are uniformly spaced between 0 and π. In this instance in particular, the condition contained in Eq. (8) is equivalent to the constraint that the magnitude of the $2M \times 2N$ point two-dimensional discrete Fourier transforms of $x(m, n)$ and $y(m, n)$ are equal.

3. RECURSIVE PHASE RETRIEVAL

As was stated in Section 2, there exists a rich and useful class of two-dimensional sequences that are uniquely defined to within some trivial ambiguities by the magnitudes of their Fourier transforms, e.g., the class of two-dimensional sequences that have finite support and irreducible z transforms. In spite of this uniqueness result, however, the reconstruction of a two-dimensional sequence from its spectral magnitude remains a difficult problem in the absence of any additional information or constraints. Therefore a number of different algorithms have been proposed that incorporate additional signal information or constraints. Gerchberg and Saxton, for example, developed an iterative algorithm that assumes, in addition to spectral magnitude, information about the magnitude of the sequence $x(m, n)$, which was assumed to be a complex-valued function of m and n.[5] Fienup, on the other hand, has considered an iterative algorithm that incorporates, in addition to a finite support constraint, a positivity constraint on $x(m, n)$.[6] As yet another example, Hayes[7] and Van Hove *et al.*[8] have investigated iterative phase-retrieval algorithms from signed Fourier-transform magnitude, i.e., Fourier-transform magnitude along with one bit of phase information. In this section, a recursive solution to the phase-retrieval problem is developed for reconstructing a two-dimensional sequence from its two-dimensional autocorrelation function $r(m, n)$ when the boundary values of $x(m, n)$ are known. Thus this algorithm is similar to those mentioned above in that some information in addition to the Fourier-transform magnitude is assumed to be known about $x(m, n)$. In this case, the additional information that is included consists of the boundary values of $x(m, n)$.

A. Development of the Algorithm

Consider an arbitrary two-dimensional sequence $x(m, n)$ whose nonzero values are contained within the rectangular region $R(M, N)$, as shown in Fig. 1(a). For convenience, it is assumed that $R(M, N)$ is the smallest possible rectangle that contains all the nonzero values of $x(m, n)$. Therefore along each edge of $R(M, N)$ there is at least one ordered pair (m, n) for which $x(m, n)$ is nonzero. The boundary of $x(m, n)$ is therefore defined as the collection of all the points of $x(m, n)$ that lie along the edges of $R(M, N)$.

The autocorrelation of $x(m, n)$, denoted by $r(m, n)$, is given by

$$\begin{aligned} r(m, n) &= x(m, n) ** x(-m, -n) \\ &= \sum_{k=0}^{M-1} \sum_{l=0}^{N-1} x(k, l) \times (m + k, n + l), \end{aligned} \tag{9}$$

where $x ** y$ is used to denote the two-dimensional convolution of x and y. Knowledge of the squared magnitude of the Fourier transform of $x(m, n)$ is equivalent to knowledge of the autocorrelation $r(m, n)$ since they form a Fourier-transform pair. Clearly, the support of $r(m, n)$ is contained within the rectangular region defined by $[-M + 1, M - 1] \times [-N + 1, N - 1]$, as shown in Fig. 1(b). Furthermore, since $x(m, n)$ is real, $r(m, n)$ is symmetric about the origin, i.e., $r(m, n) = r(-m, -n)$.

In addition, note that

$$\begin{aligned} r(m, N - 1) &= \sum_{k=0}^{M-1} x(k, 0) \times (m + k, N - 1) \\ &= x(m, 0) * x(-m, N - 1) \end{aligned} \tag{10a}$$

and

$$\begin{aligned} r(M - 1, n) &= \sum_{l=0}^{N-1} x(0, l) \times (M - 1, n + l) \\ &= x(0, n) * x(M - 1, n), \end{aligned} \tag{10b}$$

where $x * y$ is used to denote the one-dimensional convolution of x and y. With the boundary of $r(m, n)$ defined as the collection of all the points of $r(m, n)$ that lie along the edges of its region of support, note that Eqs. (10) assert that the boundary values of $r(m, n)$ may be determined from the boundary values of the sequence $x(m, n)$. The recovery of the boundary values of $x(m, n)$ from the boundary values of $r(m, n)$, however, is a nonlinear problem that, in the absence of any additional information, may not have a unique solution. Suppose, however, that the boundary values of $x(m, n)$ are known [the determination of the boundary values from $r(m, n)$ is addressed in Section 4]. More specifically, for $k = 0, 1, \ldots, N - 1$, let

$$x_k(m) = x(m, k) \quad \text{for } m = 0, 1, \ldots, M - 1 \tag{11}$$

be used to denote the one-dimensional sequence that corresponds to the kth row of the two-dimensional sequence $x(m, n)$ as shown in Fig. 1. The boundary values of $x(m, n)$ thus include the first and the last rows of $x(m, n)$, which are denoted by

$$\alpha(m) = x_0(m), \qquad \beta(m) = x_{N-1}(m), \tag{12}$$

as well as the first and the last columns of $x(m, n)$, which correspond to the first and last values of each sequence $x_k(m)$, i.e., $x_k(0)$ and $x_k(M - 1)$. Now, with $r_k(m) = r(m, k)$ used to denote the kth row of the autocorrelation sequence, as shown in Fig. 1(b), note that

$$\sum_{k=0}^{M-1} x_{N-2}(k)\alpha(m + k) + \sum_{k=1}^{M-1} \beta(k)x_1(m + k) = r_{N-2}(m) \tag{13a}$$

or

$$x_{N-2}(m) * \alpha(-m) + \beta(m) * x_1(-m) = r_{N-2}(m). \tag{13b}$$

(Recall that $*$ denotes convolution.) Therefore, with $\alpha(m)$, $\beta(m)$, and $r_{N-2}(m)$ known, Eqs. (13) represent a set of 2M − 1 *linear* equations in the unknowns $x_1(m)$ and $x_{N-2}(m)$, i.e.,

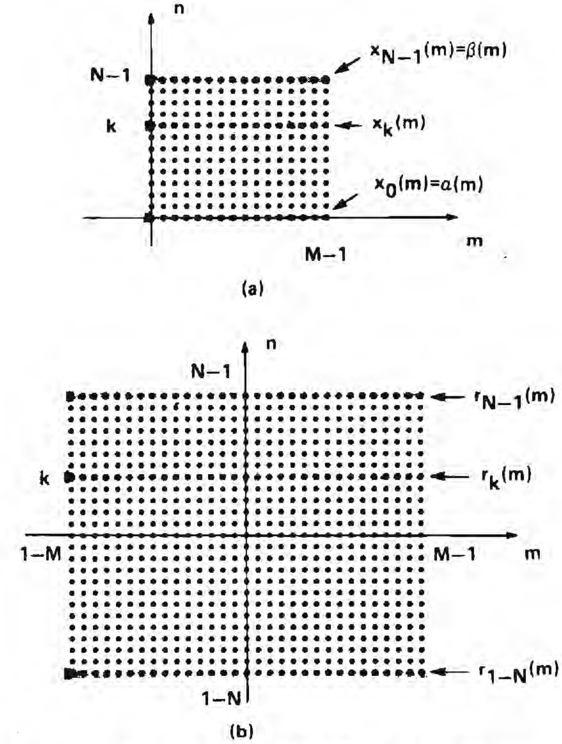


Fig. 1. (a) A region of support, $R(M, N)$, for a two-dimensional sequence. (b) The region of support of its autocorrelation.

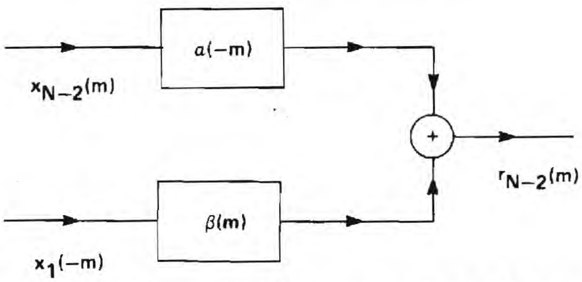


Fig. 2. System interpretation of the linear equations that define the recursive phase-retrieval algorithm.

the values of $x(m, n)$ in rows 1 and $N - 2$. A system interpretation of the set of linear equations given by Eqs. (13) is shown in Fig. 2. Specifically, Eqs. (13) define the sequence $r_{N-2}(m)$ as the sum of the outputs of two linear shift-invariant (one-dimensional) systems with unit sample responses $\alpha(-m)$ and $\beta(m)$ that are driven by the inputs $x_{N-2}(m)$ and $x_1(-m)$, respectively. The goal is to recover the unknown values of the signals $x_{N-2}(m)$ and $x_1(-m)$ from the available known information, i.e., from the signal $r(m, n)$ and the boundary values of $x(m, n)$. Recall, however, that the boundary values of $x(m, n)$ include the first and the last rows of $x(m, n)$, which correspond to the unit sample responses of the two filters in Fig. 2, as well as the first and the last columns of $x(m, n)$, which define the initial and the final values of the inputs to these filters, i.e., $x_1(0)$, $x_1(M - 1)$, $x_{N-2}(0)$, and $x_{N-2}(M - 1)$.

In order to investigate the solution to Eqs. (13), let us introduce the vector notation

$$\mathbf{x}_n = [x_n(0), x_n(1), \ldots, x_n(M-1)], \tag{14a}$$

$$\mathbf{r}_n = [r_n(1-M), r_n(2-M), \ldots, r_n(M-1)]. \tag{14b}$$

Thus Eqs. (13) may be written in matrix form as

$$\left[A \,\vdots\, B\right]\begin{bmatrix}\mathbf{x}_{N-2}\\ \hline \mathbf{x}_1\end{bmatrix} = \left[\mathbf{r}_{N-2}\right], \tag{15}$$

where A and B are $(2M - 1) \times M$ convolution matrices. As an example, for a sequence with support $R(3, 3)$, i.e., $x(m, n)$ is a 4×4 array of numbers, Eq. (15) is given by

$$\left[\begin{array}{cccc|cccc}
0 & 0 & 0 & \alpha_0 & \beta_3 & 0 & 0 & 0\\
0 & 0 & \alpha_0 & \alpha_1 & \beta_2 & \beta_3 & 0 & 0\\
0 & \alpha_0 & \alpha_1 & \alpha_2 & \beta_1 & \beta_2 & \beta_3 & 0\\
\alpha_0 & \alpha_1 & \alpha_2 & \alpha_3 & \beta_0 & \beta_1 & \beta_2 & \beta_3\\
\alpha_1 & \alpha_2 & \alpha_3 & 0 & 0 & \beta_0 & \beta_1 & \beta_2\\
\alpha_2 & \alpha_3 & 0 & 0 & 0 & 0 & \beta_0 & \beta_1\\
\alpha_3 & 0 & 0 & 0 & 0 & 0 & 0 & \beta_0
\end{array}\right]
\begin{bmatrix} x_2(0)\\ x_2(1)\\ x_2(2)\\ x_2(3)\\ x_1(0)\\ x_1(1)\\ x_1(2)\\ x_1(3)\end{bmatrix}
=
\begin{bmatrix} r_2(-3)\\ r_2(-2)\\ r_2(-1)\\ r_2(0)\\ r_2(1)\\ r_2(2)\\ r_2(3)\end{bmatrix}. \tag{16}$$

Note that $x_1(0)$, $x_1(3)$, $x_2(0)$, and $x_2(3)$ are boundary values of $x(m, n)$ and thus are assumed to be known. Therefore Eq. (16) represents seven linear equations in four unknowns. In the general case, there are $2M$ coefficients in Eq. (15) that are required in order to specify the vectors $\mathbf{x}_1$ and $\mathbf{x}_{N-2}$. The boundary values of $x(m, n)$, however, define the initial and the final values of these vectors. Consequently, Eq. (15) represents $2M - 1$ linear equations in $2M - 4$ unknowns. For the moment, it is assumed that these equations may be uniquely solved for $\mathbf{x}_1$ and $\mathbf{x}_{N-2}$. Thus, including $x_0(m) = \alpha(m)$ and $x_{N-1}(m) = \beta(m)$, the first two rows and the last two rows of $x(m, n)$ are now specified.

Now suppose that the first $(k - 1)$ rows and the last $(k - 1)$ rows of $x(m, n)$ are known, i.e., $\alpha(m)$ and $\beta(m)$ along with $x_l(m)$ and $x_{N-1-l}(m)$ for $l = 1, 2, \ldots, k - 2$. Then, as in Eqs. (13), a set of linear equations defines the unknown values in the sequnces $x_{k-1}(m)$ and $x_{N-k}(m)$. Specifically,

$$r_{N-k}(m) = x_{N-k}(m) * \alpha(-m) + \beta(m) * x_{k-1}(-m) + \sum_{l=1}^{k-2} x_{N-k+l}(m) * x_l(-m), \tag{17}$$

which may be rewritten as

$$x_{N-k}(m) * \alpha(-m) + \beta(m) * x_{k-1}(-m) = \tilde{r}_{N-k}(m), \tag{18}$$

where

$$\tilde{r}_{N-k}(m) = r_{N-k}(m) - \sum_{l=1}^{k-2} x_{N-k+l}(m) * x_l(-m) \tag{19}$$

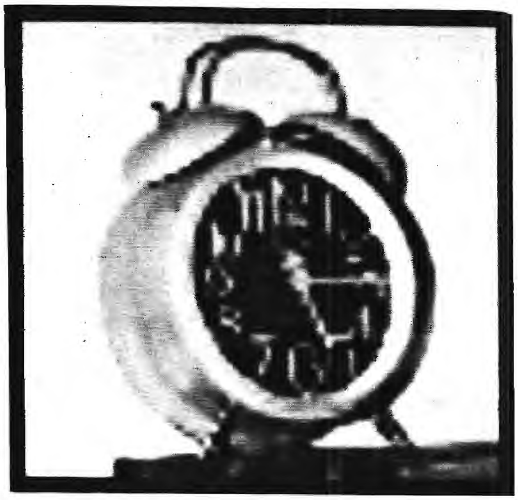
is a vector consisting of known autocorrelation values $r_{N-k}(m)$ and sums of correlations of previously computed rows of $x(m, n)$. In matrix form, Eq. (18) becomes

$$\left[A \,\vdots\, B\right]\begin{bmatrix}\mathbf{x}_{N-k}\\ \hline \mathbf{x}_{k-1}\end{bmatrix} = \left[\tilde{\mathbf{r}}_{N-k}\right], \tag{20}$$

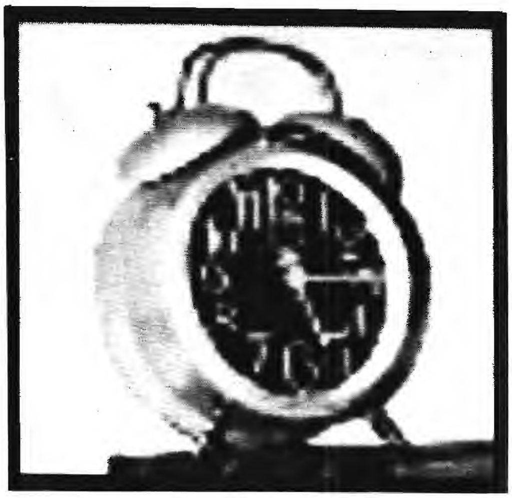
where the matrices A and B are identical to those in Eq. (15). Thus Eq. (20) provides a recursion for computing the rows $\mathbf{x}_{k-1}$ and $\mathbf{x}_{N-k}$ from the values of $\mathbf{x}_l$ and $\mathbf{x}_{N-k+l}$ for $l = 1, 2, \ldots, k - 2$. The initial conditions required to begin the recursion are the first and the last rows of $x(m, n)$, i.e., $\alpha(m)$ and $\beta(m)$. Therefore, given the boundary values of $x(m, n)$, the entire two-dimensional sequence may be recovered from its autocorrelation function by using the linear recursion [Eq. (20)], provided that the linear equations may be uniquely solved for the unknown rows. It may be shown, however, that a sufficient condition for a unique solution to Eq. (20) to exist is that $\alpha(m)$ and $\beta(m)$ not be identically zero and that $\alpha(m)$ not be related to $\beta(M - 1 - m)$ by a constant scale factor. In this case, the unknowns in Eq. (20) may be recovered by a pseudoinverse matrix operation. One interesting feature about the recursion that should be pointed out is that it requires the computation of only one pseudoinverse matrix. The recursive solution for each row consists simply of the computation of the vector $\tilde{\mathbf{r}}_{N-k}$ in Eq. (20), which is then multiplied by the pseudoinverse matrix.

B. An Example

An example that illustrates the recursive reconstruction of a two-dimensional sequence from its autocorrelation function and its boundary values is shown in Fig. 3. In particular, an



(a)



(b)

Fig. 3. Phase retrieval using known boundary conditions. (a) Original image. (b) Reconstructed image.

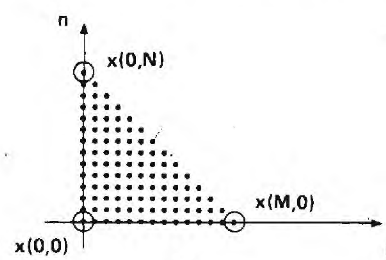
original two-dimensional sequence that has a rectangular region of support of extent 64 pixels by 64 pixels is shown in Fig. 3(a). The sequence that is obtained from the recursion [Eq. (20)] by using double-precision arithmetic is shown in Fig. 3(b) and is indistinguishable from the original. Although the recursive phase-retrieval algorithm successfully reconstructed the two-dimensional sequence in this example, this is not always the case. In particular, although it has been observed that the recursion is well suited for reconstructing two-dimensional sequences that have small regions of support, e.g., $R(M, N)$ with $M < 64$ and $N < 64$, because of the recursive nature of the algorithm the reconstruction is quite sensitive to errors that arise from computational noise. Specifically, whereas the reconstruction of large two-dimensional sequences is accurate in the initial stages of the recursion, the propagation of computational noise through the recursion decreases the accuracy of the reconstruction as the recursion progresses. Nevertheless, in reconstructing a two-dimensional sequence that has a large region of support, it is possible to consider using the recursion to reconstruct a small number of rows and columns (which may be done with a high degree of accuracy) and then use an iterative procedure in the style of Gerchberg and Saxton that, in the spatial domain, incorporates the known boundary values and the recursively computed rows and columns.

4. COMPUTATION OF THE BOUNDARY CONDITIONS

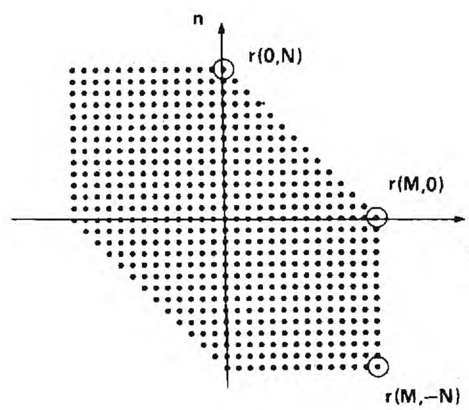
As was noted in Section 3.A, the boundary values of $x(m, n)$ are related to the boundary values of the autocorrelation function $r(m, n)$ through a set of nonlinear equations [Eqs. (10)]. Although it has been demonstrated that the solution to these equations is not necessarily unique, there are cases for which the solution is unique and for which the boundary values of $x(m, n)$ may easily be determined. Consider, for example, a two-dimensional sequence $x(m, n)$ that is known to have a triangular region of support, as shown in Fig. 4. The region of support of the autocorrelation function of $x(m, n)$ is also shown in Fig. 4. Note that the three corner points of $x(m, n)$ are related to one another by the following three second-order equations:

$$\begin{aligned} r(M, 0 &= x(M, 0)x(0, 0), \\ r(0, N) &= x(0, N)x(0, 0), \\ r(M, -N) &= x(0, N)x(M, 0). \end{aligned} \tag{21}$$

By assuming that $x(0, 0)$, $x(M, 0)$, and $x(0, N)$ are nonzero, the solution to Eqs. (21) is easily shown to be unique to within a sign. Furthermore, once these corner points are found, the entire boundary of $x(m, n)$ may easily be recovered since the boundary values of $x(m, n)$ are proportional to the boundary values of $r(m, n)$ e.g., $x(m, 0) = r(m, -N)/x(0, N)$ for $m = 0, 1, \ldots, M$. Therefore two-dimensional sequences that are known to have a triangular region of support may be easily



(a)



(b)

Fig. 4. (a) A triangular region of support for a two-dimensional sequence. (b) The region of support of its autocorrelation.

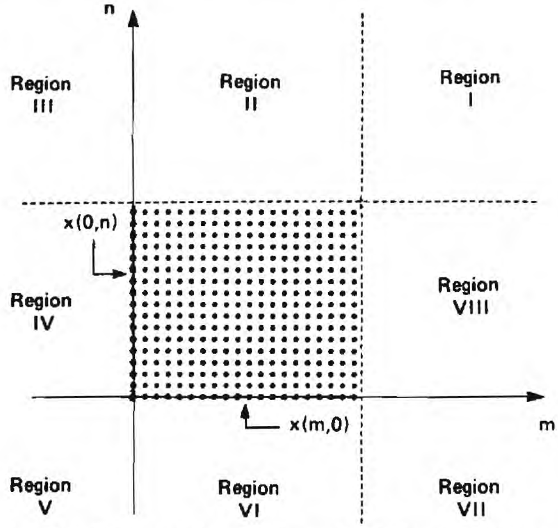


Fig. 5. The division of the two-dimensional plane into eight regions and the rectangular region $R(M, N)$.

reconstructed from only their autocorrelation, provided that the amplitudes of the corner points are nonzero.

Sequences with a triangular region of support, however, are a special case of a more-general class of sequences for which the boundary, and hence the entire sequence, may be reconstructed from its autocorrelation. In particular, consider a two-dimensional sequence that has a rectangular region of support $R(M, N)$ and suppose that the remaining two-dimensional plane is divided into the eight regions shown in Fig. 5. It is well known from off-axis holographic techniques that the incorporation of a point source sufficiently far removed from the region of support of $x(m, n)$ will allow $x(m, n)$ to be reconstructed to within a scale factor from its autocorrelation.[2] In particular, if $p(m, n) = \delta(m - k, n - l)$ is a point source at $m = k, n = l$, and if $k \geqslant 2M - 1$ or if $l \geqslant 2N - 1$, then $x(m, n)$ may be trivially reconstructed from $r(m, n)$. It is not necessary, however, that the point source $p(m, n)$ satisfy this separation constraint. Suppose, for example, that $p(m, n)$ is a unit modulus point source that lies somewhere within region I, III, V, or VII. To be more specific, let us assume that $p(m, n)$ lies in region I and, in particular, that $p(m, n) = \delta(m - M, n - N)$, as shown in Fig. 6(a). In this case, the two edges $x(0, n)$ and $x(m, 0)$ of $x(m, n)$ (illustrated in Fig. 5 by the shaded region) are easily recovered from $r(m, n)$. In particular, note that

$$r(M, N - n) = x(0, n) \qquad \text{for } n = 0, 1, \ldots, N - 1 \tag{22a}$$

and

$$r(M - m, N) = x(m, 0) \qquad \text{for } m = 0, 1, \ldots, M - 1. \tag{22b}$$

Therefore both of these edges of $x(m, n)$ correspond to the edges of the autocorrelation sequence $r(m, n)$. With these edges of $x(m, n)$ determined, it then follows from Eqs. (10) that the remaining boundary values of $x(m, n)$ may be found by a simple deconvolution or an inverse filtering operation. Therefore, if a point source of known amplitude is situated anywhere within one of the four quarter-planes defined by region I, III, V, or VII, it follows that the boundary of $x(m, n)$ may be uniquely determined from the autocorrelation $r(m, n)$, and therefore it follows from the results of Section 3.A that $x(m, n)$ may be recursively reconstructed from $r(m, n)$. It is interesting to note that, for the case in which the point source is situated at (M, N), Fiddy *et al.*[13] have shown that the z transform of the two-dimensional sequence (including the point source) is an irreducible polynomial provided that $x(M - 1, 0)$ is nonzero. Therefore, according to Theorems 1 and 2, a unique solution is guaranteed. Note also that in this case the amplitude of the point source $p(m, n)$ need not be known. Specifically, as in Eqs. (21), $p(m, n)$ and the three remaining corner points of $x(m, n)$ are related by a set of four second-order equations that may be uniquely solved for the unknowns, provided that they are nonzero.

Consider now the case in which a point source lies in region II, IV, VI, or VIII. Unlike the case described above, it is not, in general, possible to recover the boundary of $x(m, n)$ from the autocorrelation sequence $r(m, n)$. For example, consider a unit modulus point source in region II situated at (m_0, N).

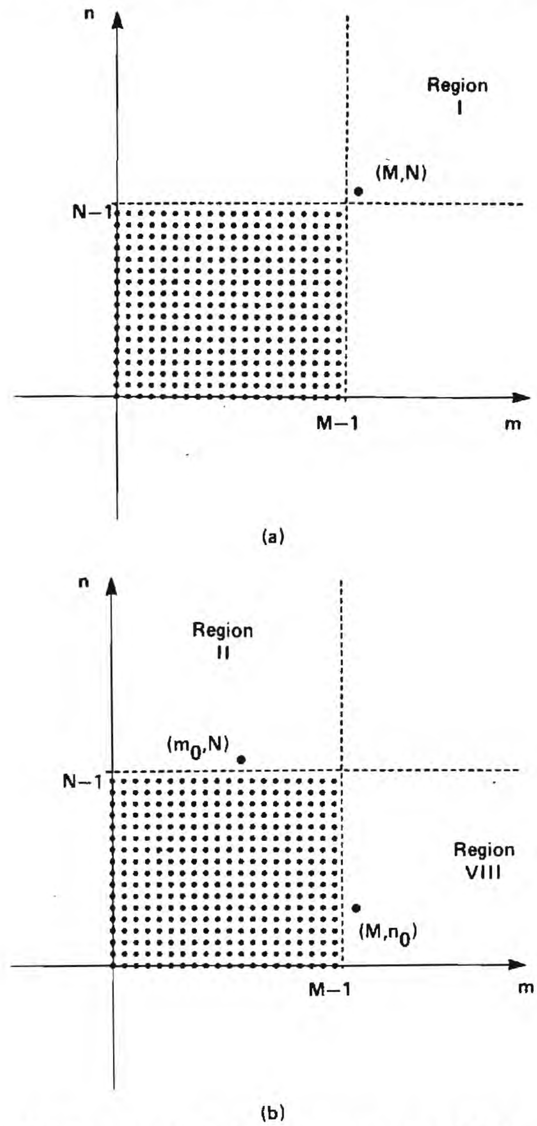


Fig. 6. Point sources sufficient for the determination of the boundary values of a two-dimensional sequence, $x(m, n)$. (a) A single point source in region I at (M, N). (b) Two point sources, one in region II at (m_0, N) and one in region VIII at (M, n_0).

In this case, the first row of $x(m, n)$ is easily derived from $r(m, n)$ since

$$r(m_0 - m, N) = x(m, 0) \quad \text{for } m = 0, 1, \ldots, M - 1. \tag{23}$$

However, without any additional information, no other boundary values of $x(m, n)$ may be determined. If, on the other hand, there are two point sources, one in region II and one in region VIII, then the complete boundary may be recovered from $r(m, n)$. More specifically, consider the two point sources shown in Fig. 6(b) that are located at (m_0, N) and (M, n_0). If these point sources have known intensities, then it follows that the first row and the first column of $x(m, n)$ may be found from $r(m, n)$. For example, if the point sources are of unit modulus, then

$$r(m_0 - m, N) = x(m, 0) \quad \text{for } m = 0, 1, \ldots, M - 1 \tag{24a}$$

and

$$r(M, n_0 - n) = x(0, n) \quad \text{for } n = 0, 1, \ldots, N - 1. \tag{24b}$$

Note that the case in which $x(m, n)$ has a triangular region of support corresponds to the case $m_0 = n_0 = 0$ above for which one point source is located in region II at $(0, N)$ and where one point source is located in region VIII at $(M, 0)$.

Although it appears that, by adding point sources, we have deviated from the problem originally addressed in Section 3, the addition of point sources is, in reality, an indirect way of defining a set of known boundary conditions. To be more specific, let $x(m, n)$ be a sequence that has a region of support given by $R(M, N)$ and assume that this is the smallest rectangle that will enclose all the nonzero values of $x(m, n)$. It follows from Eqs. (10) that the information necessary to derive the initial conditions to begin the recursive phase-retrieval algorithm are the values of $x(m, n)$ along any two contiguous edges of $R(M, N)$. Note, however, that the situations considered in Fig. 6 provide the information necessary to specify these two edges. In particular, Fig. 6(a) corresponds to the case in which all the values of $x(m, n)$ along two of the edges of its region of support are zero except for one point, i.e., the nonzero point at (M, N). In Fig. 6(b), on the other hand, the values of $x(m, n)$ along two edges are known to be zero except for the two point sources that are assumed to have known intensities.

5. SUMMARY

In this paper the importance of the boundary values of a two-dimensional sequence in the phase-retrieval problem has been investigated. Specifically, it was shown that, given the boundary values, phase retrieval becomes a linear problem that is amenable to a simple recursive solution. Furthermore, although the determination of the boundary values from only Fourier-transform magnitude information is, in general, a nontrivial problem, it was shown that, for regions of support that have certain geometries, the boundary values may easily be found. These geometries, in fact, represent a generalization of the conditions necessary for off-axis holography. An example illustrating the recursive phase-retrieval algorithm was presented, and the issue of the numerical stability of the recursion was briefly discussed.

ACKNOWLEDGMENTS

This research was supported by the National Science Foundation under grant ECS-8204793 and the Joint Services Electronics Program under contract DAAG29-81-K-0024.

REFERENCES

1. H. A. Ferwerda, "The phase reconstruction problem for wave amplitudes and coherence functions," in *Inverse Source Problems in Optics*, H. P. Baltes, ed. (Springer-Verlag, Berlin, 1978), Chap. 2.
2. L. S. Taylor, "The phase retrieval problem," IEEE Trans. Antennas Propag. **AP-29**, 386–391 (1981).
3. M. H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its Fourier transform," IEEE Trans. Acoust. Speech Signal Process. **ASSP-30**, 140–154 (1982).
4. M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," IEEE Trans. Acoust. Speech Signal Process. **ASSP-28**, 672–680 (1980).
5. R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," Optik **35**, 237–246 (1972).
6. J. R. Fienup, "Space object imaging through the turbulent atmosphere," Opt. Eng. **18**, 529–534 (1979).
7. M. H. Hayes, "The representation of signals in terms of spectral amplitude," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing* (Institute of Electrical and Electronics Engineers, New York, 1983), pp. 1446–1449.
8. P. L. Van Hove, M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from signed Fourier transform magnitude," IEEE Trans. Acoust. Speech Signal Process. (to be published).
9. M. H. Hayes and T. F. Quatieri, "The importance of boundary conditions in the phase retrieval problem," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (Institute of Electrical and Electronics Engineers, New York, 1982), pp. 1545–1548.
10. A. Mostowski and M. Stark, *Introduction to Higher Algebra* (Pergamon, New York, 1964).
11. E. M. Hofstetter, "Construction of time-limited functions with specified autocorrelation functions," IEEE Trans. Inf. Theory **IT-10**, 119–126 (1964).
12. M. H. Hayes and J. H. McClellan, "Reducible polynomials in more than one variable," Proc. IEEE **70**, 197–198 (1982).
13. M. A. Fiddy, B. J. Brames, and J. C. Dainty, "Enforcing irreducibility for phase retrieval in two dimensions," Opt. Lett. **8**, 96–98 (1983).

Signal Reconstruction from Signed Fourier Transform Magnitude

PATRICK L. VAN HOVE, MONSON H. HAYES, MEMBER, IEEE, JAE S. LIM, MEMBER, IEEE, AND ALAN V. OPPENHEIM, FELLOW, IEEE

***Abstract*—In this paper, we show that a one-dimensional or multidimensional sequence is uniquely specified under mild restrictions by its signed Fourier transform magnitude (magnitude and 1 bit of phase information). In addition, we develop a numerical algorithm to reconstruct a one-dimensional or multidimensional sequence from its Fourier transform magnitude. Reconstruction examples obtained using this algorithm are also provided.**

I. Introduction

In a variety of contexts, such as electron microscopy [1], X-ray crystallography [2], optics [3], and Fourier transform signal coding [4], it is desirable to reconstruct a sequence from partial Fourier domain information. As a consequence, considerable attention has been paid to this area, and some significant results have been developed. It has been previously established [5]-[7] that under very mild restrictions a finite extent one-dimensional (1-D) or multidimensional (MD) sequence is uniquely specified to within a scale factor by the tangent of its Fourier transform (FT) phase, and algorithms for implementing the reconstruction have been developed. It is well known that, in contrast, the FT magnitude does not uniquely specify a 1-D sequence. For MD sequences, the FT magnitude specifies a sequence to within a translation, sign, and a central symmetry [7], [8], and reconstruction algorithms developed so far have been successful [7] for only a very restricted class of MD sequences.

From the above results, on the question of unique specification of a sequence, there appear to be significant differences between 1-D and MD sequences, and between the tangent of the FT phase and the FT magnitude. In addition, the tangent of the phase and the magnitude of a complex number, which have been considered in previous studies, do not completely specify the complex number. In this paper, we show that if the signed FT magnitude (magnitude and one bit of phase information) is considered rather than the FT magnitude, there are only minor differences on the question of unique specification of a sequence, between 1-D and MD sequences, and between the tangent of the FT phase and the signed FT magnitude. In particular, it is shown that under very mild restrictions, the signed FT magnitude is sufficient to uniquely specify a 1-D or MD sequence. We note that the tangent of the phase and the signed magnitude of a complex number completely specify the complex number.

In Section II of this paper, the basic theory is presented. In Section III an algorithm for implementing the reconstruction is discussed, and Section IV illustrates several examples.

Manuscript received August 24, 1982; revised March 23, 1983. This work was supported by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grants ECS80-07102 and ECS82-04793.

P. L. Van Hove, J. S. Lim, and A. V. Oppenheim are with the Department of Electrical Engineering and Computer Science, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139.

M. H. Hayes was with the Department of Electrical Engineering and Computer Science, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139. He is now with the School of Electrical Engineering, Georgia Institute of Technology, Atlanta, GA 30332.

II. Theory

In this section, we discuss the unique specification of a sequence by its FT magnitude and 1 bit of phase. We initially consider the one-dimensional (1-D) case and then extend the 1-D result to the multidimensional (MD) case. Before we present the theoretical results, we define the notation that will be used throughout the paper.

Let $x(n)$ denote a 1-D sequence which is *causal* and *finite extent* so that $x(n)$ is zero outside $0 \leqslant n \leqslant L - 1$. Furthermore, we restrict $x(n)$ to be real-valued. Let $X(z)$ and $X(\omega)$ represent the z transform and Fourier transform of $x(n)$, so that

$$X(z) = \sum_{n=0}^{L-1} x(n) z^{-n} \tag{1}$$

$$X(\omega) = X(z) \Big|_{z=e^{j\omega}} = \sum_{n=0}^{L-1} x(n) e^{-j\omega n}. \tag{2}$$

The Fourier transform $X(\omega)$ can be represented in terms of its real part $X_R(\omega)$ and imaginary part $X_I(\omega)$, or in terms of its magnitude $|X(\omega)|$ and phase $\theta_x(\omega)$ as follows:

$$X(\omega) = X_R(\omega) + jX_I(\omega) = |X(\omega)| e^{j\theta_x(\omega)}. \tag{3}$$

To ensure that $\theta_x(\omega)$ is well defined at all ω, we assume that $X(z)$ has no zeros on the unit circle. The phase function $\theta_x(\omega)$ in (3) represents the principal value of the phase so that

$$-\pi < \theta_x(\omega) \leqslant \pi. \tag{4}$$

The 1-bit FT phase information will be represented by the function $S_x^\alpha(\omega)$ defined as

$$S_x^\alpha(\omega) = \begin{cases} +1 & \alpha - \pi \leqslant \theta_x(\omega) \leqslant \alpha \\ -1 & \text{otherwise} \end{cases} \tag{5}$$

0096-3518/83/1000-1286$01.00 © 1983 IEEE

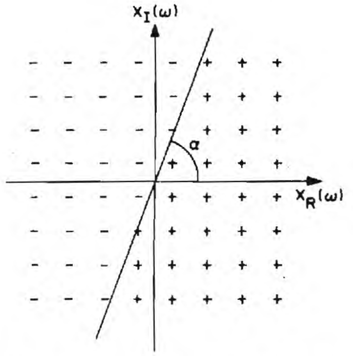


Fig. 1. Mapping of the 1-bit phase function.

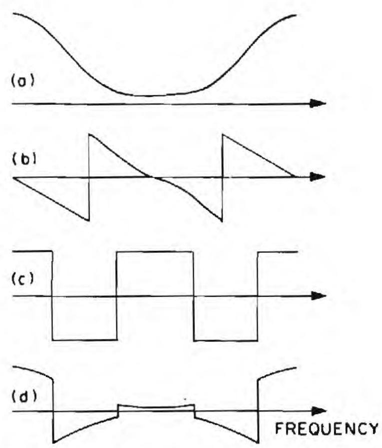


Fig. 2. Fourier transform magnitude, phase, 1-bit phase, and signed magnitude of the sequence $X(z) = 1 + 3z^{-1} + 5z^{-2} + 2z^{-3}$.

where α is a known constant in the range of $0 < \alpha \leqslant \pi$. Thus, the complex plane is divided into two regions separated by a straight line passing through the origin and at an angle α with the real axis, as shown in Fig. 1. For example, for $\alpha = \pi/2$, $S_x^{\pi/2}(\omega)$ represents the algebraic sign of $\mathrm{Re}\{X(\omega)\}$. More generally, $S_x^{\alpha}(\omega)$ is the algebraic sign of $\mathrm{Re}\{e^{j(\pi/2-\alpha)}X(\omega)\}$. The algebraic sign of zero is assumed to be positive.

The function $G_x^{\alpha}(\omega)$ is defined as

$$G_x^{\alpha}(\omega) = S_x^{\alpha}(\omega)|X(\omega)| \tag{6}$$

and will be referred to as the signed Fourier transform magnitude since it contains both magnitude and sign information. An example of $|X(\omega)|$, $\theta_x(\omega)$, $S_x^{\alpha}(\omega)$, and $G_x^{\alpha}(\omega)$ when $\alpha = \pi/2$ and $X(z) = 1 + 3z^{-1} + 5z^{-2} + 2z^{-3}$ is shown in Fig. 2.

Finally, given a positive integer N, we define a constant P and an interval R as

$$P = \frac{N-1}{2} \text{ and } R = (0, \pi) \text{ for } N \text{ odd}$$

$$P = \frac{N}{2} \text{ and } R = (0, \pi] \text{ for } N \text{ even.} \tag{7}$$

The uniqueness of a 1-D sequence when the signed Fourier transform magnitude $G_x^{\alpha}(\omega)$ is specified is based on the following statements. The proof of these statements is given in the Appendix.

Statement A1: Let $x(n)$ and $y(n)$ be two real, causal, and finite extent sequences. If $|X(\omega)| = |Y(\omega)|$, $x(n)$ and $y(n)$ can always be expressed as

$$x(n) = b(n) * a(n)$$

and

$$y(n) = \epsilon b(n) * a(N-1-n)$$

where $\epsilon = +1$ or -1 and $a(n)$ and $b(n)$ are real, causal, and finite extent sequences with N corresponding to the length of $a(n)$, i.e., $a(n) = 0$ outside $0 \leqslant n \leqslant N-1$.

Statement A2: Let $b(n)$ be a real, causal, and finite extent sequence. For any positive integer N, the equation

$$\mathrm{Re}\{B(z)\, z^{-(N-1)/2}|_{z=e^{j\omega}}\} = 0$$

is satisfied for at least P distinct values of ω in the interval R, where P and R are as defined in (7).

Statement A3: Let $a(n)$ be a real sequence which is zero outside $0 \leqslant n \leqslant N-1$. If the equation

$$\mathrm{Im}\{A(z)\, z^{(N-1)}|_{z=e^{j\omega}}\} = 0$$

is satisfied for at least P distinct values of ω in the interval R, then it is identically equal to zero and $a(n) = a(N-1-n)$.

We use the above three statements, whose proofs are shown in the Appendix, to demonstrate the following theorem:

Theorem 1: Let $x(n)$ and $y(n)$ be two real, causal, and finite extent sequences with z transforms which have no zeros on the unit circle. If $G_x^{\pi/2}(\omega) = G_y^{\pi/2}(\omega)$ for all ω, then $x(n) = y(n)$.

To show Theorem 1, we note from (5) and (6) that the condition $G_x^{\pi/2}(\omega) = G_y^{\pi/2}(\omega)$ is equivalent to

$$\mathrm{sign}\{X_R(\omega)\}\,|X(\omega)| = \mathrm{sign}\{Y_R(\omega)\}\,|Y(\omega)| \tag{8}$$

which in turn implies that $|X(\omega)| = |Y(\omega)|$, and therefore that

$$\mathrm{sign}\{X_R(\omega)\} = \mathrm{sign}\{Y_R(\omega)\}. \tag{9}$$

From Statement A1, then, $x(n)$ and $y(n)$ can be expressed as

$$x(n) = b(n) * a(n)$$

$$y(n) = \epsilon\, b(n) * a(N-1-n) \tag{10}$$

where $\epsilon = \pm 1$. Fourier transforming (10) we obtain

$$X(\omega) = A(\omega)\, B(\omega)$$

$$Y(\omega) = \epsilon\, e^{-j\omega(N-1)}\, A(-\omega)\, B(\omega). \tag{11}$$

To show that $\epsilon = 1$ in (11), we evaluate (9) at $\omega = 0$ and recognize that $X_R(0) = A(0)\, B(0)$ and $Y_R(0) = \epsilon A(0)\, B(0)$, so that

$$\mathrm{sign}\{A(0)\, B(0)\} = \mathrm{sign}\{\epsilon A(0)\, B(0)\}. \tag{12}$$

Since $X(\omega)$ is not zero at $\omega = 0$, (12) requires that $\epsilon = +1$.

Since $\epsilon = 1$, from (10), showing that $x(n) = y(n)$ is equivalent to showing that $a(n) = a(N-1-n)$. Toward this end, we consider the sum

$$X_R(\omega) + Y_R(\omega).$$

From (11) with $\epsilon = 1$, it can be shown that

$$X_R(\omega) + Y_R(\omega) = 2\, \mathrm{Re}[A(\omega) e^{j\omega(N-1)/2}] \cdot \mathrm{Re}[B(\omega) e^{-j\omega(N-1)/2}]. \tag{13}$$

From Statement A2, there are at least P distinct values of ω in the interval R which we denote as ω_i, $i = 1, 2, \cdots, P$ for which

$$\mathrm{Re}[B(\omega_i)e^{-j\omega_i(N-1)/2}] = 0, \quad i = 1, 2, \cdots, P, \omega_i \in R. \tag{14}$$

From (13) and (14),

$$X_R(\omega_i) + Y_R(\omega_i) = 0, \quad i = 1, 2, \cdots, P, \omega_i \in R. \tag{15}$$

From (9), both terms of the left-hand side of (15) have the same sign for all ω. Since a sum of two terms having the same sign can be zero only when both terms are zero, we have

$$X_R(\omega_i) = Y_R(\omega_i) = 0$$

and therefore also,

$$X_R(\omega_i) - Y_R(\omega_i) = 0, \quad i = 1, 2, \cdots, P, \omega_i \in R. \tag{16}$$

From (11) and the fact that $\epsilon = 1$, it can be shown that (16) can be expressed as

$$\begin{aligned} X_R(\omega_i) - Y_R(\omega_i) = -2\,\mathrm{Im}[A(\omega_i)e^{j\omega_i(N-1)/2}] \\ \cdot \mathrm{Im}[B(\omega_i)e^{-j\omega_i(N-1)/2}] = 0, \\ i = 1, 2, \cdots, P, \omega_i \in R. \end{aligned} \tag{17}$$

Since $B(\omega)$ is not zero for any ω, it follows from (14) that the second factor in (17) satisfies the property

$$\mathrm{Im}[B(\omega_i)e^{-j\omega_i(N-1)/2}] \neq 0, \quad i = 1, 2, \cdots, P, \omega_i \in R. \tag{18}$$

From (17) and (18),

$$\mathrm{Im}[A(\omega_i)e^{j\omega_i(N-1)/2}] = 0, \quad i = 1, 2, \cdots, P, \omega_i \in R. \tag{19}$$

From (19) and Statement A3, $a(n) = a(N - 1 - n)$ so that $x(n) = y(n)$, thus demonstrating Theorem 1.

The result in Theorem 1 can be generalized in various ways. Specifically, in Theorem 1, we have assumed that $\alpha = \pi/2$, which is a specific representation of the 1-bit phase information. It can be shown that the statement is true for other choices of $0 < \alpha < \pi$. When $\alpha = \pi$ so that $S_x^{\pi}(\omega) = \mathrm{sign}[\theta_x(\omega)]$, a sequence is uniquely specified by $G_x^{\pi}(\omega)$ when $x(0) = 0$. Theorem 1 can also be extended to anticausal (left-sided) sequences. The proofs of these extensions can be found in [9]. When the above extensions are incorporated in Theorem 1, we have the following general theorem:

Theorem 2: Let $x(n)$ and $y(n)$ be two real, causal (or anticausal), and finite extent sequences, with z transforms which have no zeros on the unit circle. If $G_x^{\alpha}(\omega) = G_y^{\alpha}(\omega)$ for all ω and $0 < \alpha < \pi$, then $x(n) = y(n)$. When $\alpha = \pi$, if $G_x^{\pi}(\omega) = G_y^{\pi}(\omega)$ and $x(0) = y(0) = 0$, then $x(n) = y(n)$.

Theorems 1 and 2 explicitly require that the sequences be real-values and causal (or anticausal). The necessity of these conditions can be illustrated through counterexamples. Consider first the condition that the sequences be real, and let $y(n)$ equal $e^{j(\alpha-\pi)}x(n)$ where $x(n)$ is real. In this case, it is straightforward to show that $G_x^{\pi}(\omega) = G_y^{\alpha}(\omega)$. Since $G_x^{\pi}(\omega)$ does not uniquely specify $x(n)$, $G_y^{\alpha}(\omega)$ does not uniquely specify $y(n)$. To indicate the necessity of the causality (or anticausality) condition, consider as one counterexample the two-sided sequences $x(n)$ and $y(n)$ for which the z transforms are

$$\begin{aligned} X(z) &= -z^2 + 6 - z^{-2} = (z + 2 - z^{-1})(-z + 2 + z^{-1}) \\ Y(z) &= z^2 + 4z + 2 - 4z^{-1} + z^{-2} = (z + 2 - z^{-1})^2. \end{aligned} \tag{20}$$

For these two sequences it can be easily shown that $|X(\omega)| = |Y(\omega)|$ and $S_x^{\pi/2}(\omega) = S_y^{\pi/2}(\omega)$. In this case, then, $x(n)$ and $y(n)$ are different sequences, but they have the same signed FT magnitude.

In Theorems 1 and 2, uniqueness results were presented assuming that the signed spectral magnitude of a finite length sequence is known for all frequencies in the interval $(0, 2\pi)$. In the case of FT phase, it is possible to generalize the uniqueness results to the case in which the FT phase is known only for a finite number of distinct frequencies. Specifically, it has been shown [6] that for a finite length sequence of length N which has no symmetric (zero-phase) factors in its z transform, any $(N - 1)$ samples of the FT phase are sufficient to uniquely define the sequence to within a scale factor. Therefore, since the FT phase need not be known for all ω, such a result has been useful [6] in the development of practical algorithms for recontructing a finite length sequence from its FT phase samples. Unfortunately, however, a fixed finite set of signed magnitude samples is not always sufficient to uniquely specify a real, causal, and finite length sequence. For example, consider the following two causal sequences of length $N = 3$.

$$x(n) = 1.0\,\delta(n) + 2.6\,\delta(n-1) + 1.2\,\delta(n-2) \tag{21}$$

$$y(n) = 1.2\,\delta(n) + 2.6\,\delta(n-1) + 1.0\,\delta(n-2). \tag{22}$$

Since $y(n)$ is obtained from $x(n)$ by flipping both of the zeros of $X(z)$ about the unit circle, both $x(n)$ and $y(n)$ have the same spectral magnitude. Furthermore, in the interval $(0, \pi)$ the real part of the Fourier transform of $x(n)$ is equal to zero at only one frequency, $\omega = 0.477023\pi$ and the real part of the Fourier transform of $y(n)$ is equal to zero only at $\omega = 0.526166\pi$. Therefore, the signed magnitude of $X(\omega)$ is equal to the signed magnitude of $Y(\omega)$ for all ω outside the intervals $(0.477023\pi, 0.526166\pi)$ and $(-0.526166\pi, -0.477023\pi)$. Consequently, an arbitrary number of signed magnitude samples within this region is not sufficient to distinguish $x(n)$ from $y(n)$.

Even though a real, causal, finite extent sequence is not uniquely specified by samples of its signed FT magnitude at a finite number of arbitrary frequencies, it is specified by samples of its signed FT magnitude at a finite number of properly chosen frequencies which are different for different sequences. Specifically, for $x(n)$ which is zero outside $0 \leqslant n \leqslant N - 1$, the FT magnitude $|X(\omega)|$ is completely specified by $(N - 1)$ discrete Fourier transform (DFT) samples in the interval $(0, \pi)$. The 1 bit of FT phase $S_x^{\alpha}(\omega)$ is completely specified by the positions of its discontinuities and by its value at $\omega = 0$. Since the function $S_x^{\alpha}(\omega)$ has at most $2N$ discontinuities in $(-\pi, +\pi)$, $G_x^{\alpha}(\omega)$ is completely specified by a maximum of $3N$ samples at properly chosen frequencies.

In the above discussion, we considered only 1-D sequences. We now extent Theorem 2 to MD sequences. Let $x(\boldsymbol{n})$ denote an MD sequence $x(n_1, n_2, \cdots, n_M)$, and let $G_x^{\alpha}(\boldsymbol{\omega})$ denote the signed FT magnitude of $x(\boldsymbol{n})$, where $G_x^{\alpha}(\boldsymbol{\omega})$ represents $G_x^{\alpha}(\omega_1, \omega_2, \cdots, \omega_M)$ and is given by $S_x^{\alpha}(\boldsymbol{\omega})|X(\boldsymbol{\omega})|$. We define an MD sequence $x(\boldsymbol{n})$ to have a one-sided region of support in the M-dimensional space $n_1, n_2, \cdots, n_M$ if it has nonzero values for only one polarity of each n_i. For example, for a two-dimensional sequence there are four possible regions of support which are consistent with the sequence being one sided, corresponding to the four quardrants. Theorem 3, which follows, represents a generalization of Theorem 2 to encompass MD sequences.

Theorem 3: Let $x(\mathbf{n})$ and $y(\mathbf{n})$ be two real finite extent sequences with one-sided support and with z transforms which have no zeros at $|z_1| = |z_2| = \cdots = |z_M| = 1$. If $G_x^\alpha(\boldsymbol{\omega}) = G_y^\alpha(\boldsymbol{\omega})$ for all $\boldsymbol{\omega}$ and $0 < \alpha < \pi$, then $x(\mathbf{n}) = y(\mathbf{n})$. When $\alpha = \pi$, if $G_x^\pi(\boldsymbol{\omega}) = G_y^\pi(\boldsymbol{\omega})$ and $x(0) = y(0) = 0$, then $x(\mathbf{n}) = y(\mathbf{n})$.

We demonstrate the validity of Theorem 3 for a 2-D sequence which has the first-quadrant support size $M_1 \times M_2$ so that

$$x(n_1, n_2) = y(n_1, n_2) = 0 \text{ outside } 0 \leq n_1 \leq M_1 - 1 \text{ and } 0 \leq n_2 \leq M_2 - 1.$$

The proof for a higher dimension and for a different quadrant support is analogous to the 2-D case with the first-quadrant support. To demonstrate Theorem 3, we map the 2-D sequences $x(n_1, n_2)$ and $y(n_1, n_2)$ into two 1-D sequences $\hat{x}(n)$ and $\hat{y}(n)$ by the following transformation:

$$\hat{x}(n_1 \cdot M_2 + n_2) = x(n_1, n_2)$$
$$\hat{y}(n_1 \cdot M_2 + n_2) = y(n_1, n_2). \tag{23}$$

In essence, the transformation in (23) corresponds to mapping a 2-D sequence to a 1-D sequence by concatenating the columns of the 2-D sequence. Clearly, $\hat{x}(n)$ and $\hat{y}(n)$ given by (23) are real, causal, and finite extent sequences. From (23), it is clear that the transformation is invertible. Furthermore, it can be shown [10] that

$$\hat{X}(\omega) = X(\omega_1, \omega_2)\Big|_{\omega_1 = \omega \cdot M_2,\ \omega_2 = \omega}$$

and

$$\hat{Y}(\omega) = Y(\omega_1, \omega_2)\Big|_{\omega_1 = \omega \cdot M_2,\ \omega_2 = \omega}. \tag{24}$$

From (24), it follows that the signed FT magnitudes of $\hat{x}(n)$ and $\hat{y}(n)$ are specified by the signed FT magnitudes of $x(n_1, n_2)$ and $y(n_1, n_2)$. Therefore, if $G_x^\alpha(\omega_1, \omega_2) = G_y^\alpha(\omega_1, \omega_2)$, then $G_{\hat{x}}^\alpha(\omega) = G_{\hat{y}}^\alpha(\omega)$. In addition, since $X(z_1, z_2)$ and $Y(z_1, z_2)$ have no zeros at $|z_1| = |z_2| = 1$, from (24), $\hat{X}(z)$ and $\hat{Y}(z)$ have no zeros on the unit circle. Since $\hat{x}(n)$ and $\hat{y}(n)$ satisfy all the conditions in Theorem 2, it follows from Theorem 2 that $\hat{x}(n) = \hat{y}(n)$. Since the transformation (23) is invertible, $x(n_1, n_2) = y(n_1, n_2)$ as required by Theorem 3.

The condition that $X(\boldsymbol{\omega}) \neq 0$ at any $\boldsymbol{\omega}$ is much more restrictive for 2-D sequences than for 1-D sequences, since $X(z) = 0$ represents surfaces in the (z_1, z_2) plane for 2-D sequences and points in the z plane for 1-D sequences. From the proof of Theorem 3 described above, however, it is not necessary to require $X(\boldsymbol{\omega}) \neq 0$ at any $\boldsymbol{\omega}$. We only need to require that $X(\boldsymbol{\omega}) \neq 0$ at the slices of $\boldsymbol{\omega}$ needed to form $\hat{X}(\boldsymbol{\omega})$ in (24). This is a much less restrictive condition than the condition in Theorem 3.

The theoretical result in Theorem 3 differs from that by Hayes [5] in several respects. In the result by Hayes [5], only samples of the FT magnitude are required, but the sequence is restricted to have a nonfactorizable z transform and the unique specification of the sequence is only to within a sign, a translation, and a central symmetry. In Theorem 3, the signed FT magnitude is required, but the sequence may have a factorizable z transform and is uniquely specified in the strict sense.

III. Algorithm

In Section II, we showed that under certain conditions a sequence is uniquely specified by its signed FT magnitude. In this section, we discuss an algorithm to implement the reconstruction of a sequence $x(\mathbf{n})$ from its signed FT magnitude. The sequence $x(\mathbf{n})$ is assumed to satisfy the conditions of Theorem 3. In addition, its signed FT magnitude $G_x^\alpha(\boldsymbol{\omega})$ is assumed known.

The algorithm that we have developed is an iterative procedure which is similar in style to other iterative procedures studied by Gerchberg-Saxton [11] and Fienup [12]. In the iterative algorithm, the "time" domain constraint that $x(\mathbf{n})$ is real and finite extent with a one-sided region of support, and the frequency domain constraint that the signed FT magnitude of $x(\mathbf{n})$ is given by $G_x^\alpha(\boldsymbol{\omega})$, are imposed separately in each iteration. Specifically, let $X_p(\boldsymbol{\omega})$ denote the estimate of $X(\boldsymbol{\omega})$ at the pth iteration. The estimate $X_p(\boldsymbol{\omega})$ is inverse Fourier transformed to the time domain to obtain $x_p'(\mathbf{n})$

$$x_p'(\mathbf{n}) = F^{-1}[X_p(\boldsymbol{\omega})]. \tag{25}$$

From $x_p'(\mathbf{n})$, we generate an estimate $x_p''(\mathbf{n})$ which satisfies the time domain constraints

$$x_p''(\mathbf{n}) = \begin{cases} \mathrm{Re}[x_p'(\mathbf{n})] & \text{for } \mathbf{n} \in A \\ 0 & \text{for } \mathbf{n} \notin A \end{cases} \tag{26}$$

where A represents the known support region of $x(\mathbf{n})$.

The sequence $x_p''(\mathbf{n})$ is then Fourier transformed back to the frequency domain to obtain $X_p''(\boldsymbol{\omega})$ as follows:

$$X_p''(\boldsymbol{\omega}) = F[x_p''(\mathbf{n})]. \tag{27}$$

The new frequency domain estimate $X_{p+1}(\boldsymbol{\omega})$ is then obtained by enforcing the constraint that $G_{x_{p+1}}^\alpha(\boldsymbol{\omega}) = G_x^\alpha(\boldsymbol{\omega})$ as follows:

$$X_{p+1}(\boldsymbol{\omega}) = \begin{cases} |X(\boldsymbol{\omega})| e^{j\theta_{x_p''}(\boldsymbol{\omega})} & \text{if } S_{x_p''}^\alpha(\boldsymbol{\omega}) = S_x^\alpha(\boldsymbol{\omega}) \\ |X(\boldsymbol{\omega})| e^{j(2\alpha - \theta_{x_p''}(\boldsymbol{\omega}))} & \text{if } S_{x_p''}^\alpha(\boldsymbol{\omega}) = -S_x^\alpha(\boldsymbol{\omega}). \end{cases} \tag{28}$$

Specifically, the correct magnitude is substituted for the estimated magnitude. If $S_{x_p''}^\alpha(\boldsymbol{\omega}) = S_x^\alpha(\boldsymbol{\omega})$, then the phase of the estimate is retained. Otherwise, the estimate is reflected about a line that passes through the origin with angle α to correct the sign of $S_{x_p''}^\alpha(\boldsymbol{\omega})$. This completes one iteration. The initial estimate $X_0(\boldsymbol{\omega})$ we have used is given by

$$X_0(\boldsymbol{\omega}) = |X(\boldsymbol{\omega})| e^{j\theta_{x_0}(\boldsymbol{\omega})} \tag{29}$$

where $\theta_{x_0}(\boldsymbol{\omega})$ is given by

$$\theta_{x_0}(\boldsymbol{\omega}) = \begin{cases} \alpha - \dfrac{\pi}{2} & \text{for } S_x^\alpha(\boldsymbol{\omega}) = +1 \\ \alpha + \dfrac{\pi}{2} & \text{for } S_x^\alpha(\boldsymbol{\omega}) = -1. \end{cases} \tag{30}$$

The iterative algorithm discussed above is illustrated in Fig. 3.

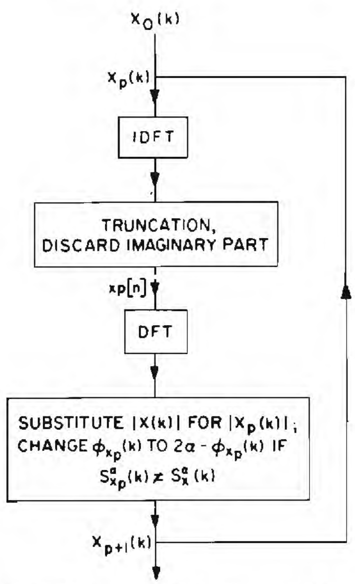


Fig. 3. Block diagram of the iterative algorithm.

The asymptotic behavior of the algorithm in Fig. 3 has not yet been studied theoretically. We have observed experimentally that a stable estimate of the sequence to be retrieved is always attained after a large number of iterations.

To implement the algorithm in Fig. 3, the Fourier and inverse Fourier transform operations are approximated by discrete Fourier transfrom (DFT) and inverse DFT (IDFT) operations. Although the uniqueness is not guaranteed in terms of the signed FT magnitude samples, we have empirically observed that the algorithm reconstructs the desired sequence provided that the signed FT magnitude is densely sampled in the frequency domain, so that the FT magnitude is completely specified and the discontinuities of $S_x^{\alpha}(\boldsymbol{\omega})$ are individually resolved by the samples of $S_x^{\alpha}(\boldsymbol{\omega})$. The FT magnitude $|X(\boldsymbol{\omega})|$ is completely specified by samples of $|X(\boldsymbol{\omega})|$ when the DFT size is twice the size of the known support of $x(\boldsymbol{n})$ in each dimension.

IV. Examples

The algorithm discussed in Section III has been used to reconstruct a variety of different 1-D and 2-D sequences from their signed FT magnitudes. In this section, we present some of these examples.

Fig. 4 illustrates one example in which a 1-D sequence is reconstructed from its signed FT magnitude. In Fig. 4(a) is shown a 47-point sequence obtained by sampling female speech at a 10 kHz rate. In Fig. 4(b) is shown the sequence reconstructed by using the iterative algorithm with the DFT size of 1024 after 50 iterations. In addition to the above example, a number of other examples have been considered. In all cases, we observed that the algorithm reconstructs the desired sequence.

Fig. 5 illustrates an example in which a 2-D sequence is reconstructed from its signed FT magnitude. In Fig. 5(a) is shown an image of size 256 X 256 pixels. In Fig. 5(b) is shown the image reconstructed by using the iterative algorithm using the DFT size of 512 X 512 after 10 iterations.

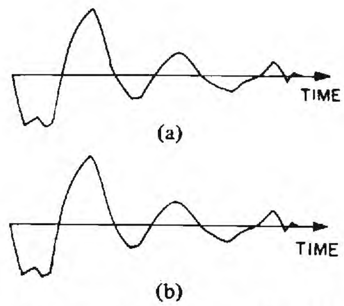


Fig. 4. Speech segment sampled at 47 points. (a) Original sequence. (b) Reconstructed sequence after 50 iterations.

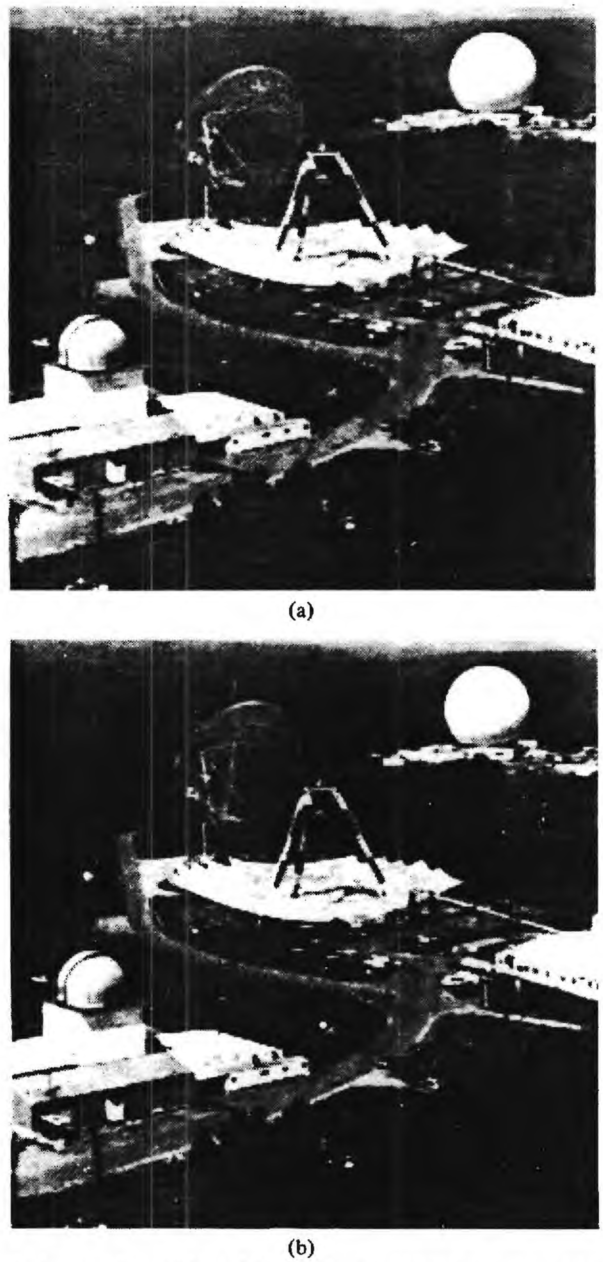


Fig. 5. Image of size 256 X 256 pixels. (a) Original image. (b) Reconstructed image after 10 iterations.

In addition to the examples shown in this section, we have studied a number of other examples. From these examples, we have made the following observations about the iterative algorithm. First, for sequences satisfying the uniqueness constraints, if a DFT size below some threshold value is used, the algorithm does not lead to the desired sequence. The threshold value is different for different sequences, and we have not yet found a simple way to determine the threshold value for a given sequence. In practice, therefore, the DFT size is typically much larger than the threshold value to reconstruct a sequence from its signed FT magnitude. Second, the DFT size required is typically much larger (by more than a factor of 10 typically) than the size of the data for 1-D signals. For 2-D signals, we have observed that the DFT size of $2N \times 2N$ when the data size is $N \times N$ is sufficient for all examples we considered. This difference is in part due to the fact that the magnitude of $2N \times 2N$ DFT when the data size is $N \times N$ uniquely specifies a 2-D sequence within a sign factor, a translation, and a central symmetry, and therefore the ambiguity that needs to be resolved by 1 bit of phase information is much less for 2-D signals than for 1-D signals. Third, the threshold DFT length is approximately the same for different choices of α, as long as α is not too close to 0 or π. As α approaches 0 or π, the thresold length is significantly increased. The choice of $\alpha = \pi/2$ permits the use of FFT routines specific to real sequences, and therefore, uses less computation time and less storage space. Fourth, the convergence rate of the iterative algorithm is rapid initially and becomes slow as the number of iterations is increased. Fifth, we have observed that the mean square error between the original and reconstructed sequences decreases monotonically as the number of iterations increases. Sixth, the convergence rate of the algorithm can be significantly improved by using an acceleration procedure similar to that used by Oppenheim *et al.* [13]. Further details on the behavior of the iterative algorithm can be found in Van Hove [9].

V. Conclusions

In this paper, we have shown that a 1-D or MD sequence is uniquely specified under mild restrictions by its signed FT magnitude. In addition, we have developed an iterative algorithm to reconstruct a 1-D or MD sequence from its signed FT magnitude. When this result is combined with the previous result [5] on the problem of reconstructing a 1-D or MD sequence from its FT phase, we obtain a very general result that a 1-D or MD sequence is uniquely specified by its FT phase or its signed FT magnitude. In addition, under mild restrictions, an iterative algorithm which is similar in style can be used to reconstruct a 1-D or MD sequence from its FT phase or signed magnitude.

Appendix

Statement A1: Let $x(n)$ and $y(n)$ be two real, causal, and finite extent sequences. If $|X(\omega)| = |Y(\omega)|$, $x(n)$ and $y(n)$ can always be expressed as

$$x(n) = b(n) * a(n)$$
$$y(n) = \epsilon b(n) * a(N - 1 - n)$$

where $\epsilon = +1$ or -1 and $a(n)$ and $b(n)$ are real, causal, and finite extent with N corresponding to the length of $a(n)$, i.e., $a(n) = 0$ outside $0 \leqslant n \leqslant N - 1$.

Proof: A general expression of the z transform $X(z)$ of a sequence $x(n)$ which is causal and has a finite support is given by

$$X(z) = z^{-n_1} x_0 \prod_{i=1}^{Q} (1 - z_i z^{-1}) \tag{A1.1}$$

where z_i, $i = 1, 2, \cdots, Q$, are the zeros of $X(z)$, x_0 is the first nonzero sample, and n_1 is the positive initial delay in $x(n)$. It is well known that the FT magnitude of a finite extent 1-D sequence remains unchanged only when the sequence is subject to linear shifts, sign inversions, and/or zero "flipping." The z transform $Y(z)$ may therefore be written as

$$Y(z) = \pm z^{-n_2} x_0 \prod_{i \in \{u\}} (1 - z_i z^{-1}) \prod_{i \in \{r\}} (-z_i + z^{-1}) \tag{A1.2}$$

where n_2 is the positive initial delay in $y(n)$, $\{r\}$ is the set of indexes of the R zeros of $Y(z)$ which are zeros of $X(z)$ reflected across the unit circle, and $\{u\}$ is the set of indexes of zeros which are unchanged from $X(z)$ to $Y(z)$. We may also write (A1.1) and (A1.2) as

$$X(z) = A(z) \cdot B(z)$$
$$Y(z) = \pm C(z) \cdot B(z)$$

or

$$x(n) = a(n) * b(n)$$
$$y(n) = \pm c(n) * b(n) \tag{A1.3}$$

where

$$A(z) = z^{-(n_1 - n_2)} \prod_{i \in \{r\}} (1 - z_i z^{-1})$$
$$B(z) = z^{-n_2} x_0 \prod_{i \in \{u\}} (1 - z_i z^{-1})$$
$$C(z) = \prod_{i \in \{r\}} (-z_i + z^{-1}). \tag{A1.4}$$

We now show that $c(n)$ is $a(n)$ time reversed, represented by $a'(n)$. The length of the sequence $a'(n)$ is $N = n_1 - n_2 + R + 1$, if we include the leading zeros. Therefore,

$$a'(n) = a(N - 1 - n)$$
$$A'(z) = A(z^{-1}) z^{-(N-1)} = z^{-R} \prod_{i \in \{r\}} (1 - z_i z) = C(z)$$

so that $c(n) = a(N - 1 - n)$. From (A1.3), the sequences $x(n)$ and $y(n)$ are expressed in the adequate form. To characterize $a(n)$ and $b(n)$, we examine their z transforms. Since $B(z)$ contains only a finite number of negative powers of z, the sequence $b(n)$ has a finite causal support. Since $A(z)$ and $A'(z) = C(z)$ contain only negative powers of z, it follows that $a(n)$ and $a(N - 1 - n)$ are causal so that $a(n)$ is zero outside $0 \leqslant n \leqslant N - 1$. If the z transform $X(z)$ contains a pair of complex conjugate zeros, than they must both belong to $\{u\}$ or both to $\{r\}$ for $y(n)$ to be real-valued. The z transforms $A(z)$ and $B(z)$ may therefore contain complex zeros only in conjugate pairs so that $a(n)$ and $b(n)$ are real. In the case $n_2 > n_1$, we simply exchange the roles of $x(n)$ and $y(n)$. This completes the proof of Statement A1.

Statement A2: Let $b(n)$ be a real, causal, and finite extent sequence. For any positive integer N, the equation

$$\operatorname{Re}\{B(z)\, z^{-(N-1)/2}|_{z=e^{j\omega}}\} = 0$$

is satisfied for at least P distinct values of ω in the interval R where P and R are as defined in (7) of the text.

To prove this statement, we introduce the notion of unwrapped phase. Given a Fourier transform $M(\omega)$ which has no zeros, we define its unwrapped phase $\phi_M(\omega)$ as the unique continuous function of ω which satisfies

$$M(\omega) = |M(\omega)| e^{j\phi_M(\omega)} \tag{A2.1}$$

for all ω and which takes the value of 0 or $-\pi$ at $\omega = 0$. The unwrapped phase has the following properties. If we define the function $F(\omega)$ as

$$F(\omega) = D(\omega)\, B(\omega) \tag{A2.2}$$

then if follows that

$$\phi_F(\omega) = \phi_D(\omega) + \phi_B(\omega) + 2\alpha\pi$$

where

$$\alpha = \begin{cases} 1 & \text{if } \phi_D(0) = \phi_B(0) = -\pi \\ 0 & \text{otherwise.} \end{cases} \tag{A2.3}$$

The unwrapped FT phase $\phi_B(\omega)$ of a causal sequence $b(n)$ satisfies

$$\phi_B(0) \geqslant \phi_B(\pi). \tag{A2.4}$$

The unwrapped phase of the function

$$D(\omega) = e^{-j\omega(N-1)/2} \tag{A2.5}$$

is

$$\phi_D(\omega) = -\omega \frac{N-1}{2}. \tag{A2.6}$$

We now proceed to the proof of statement A2. We consider the unwrapped phase $\phi_F(\omega)$ of the function

$$F(\omega) = B(\omega) e^{-j\omega(N-1)/2}.$$

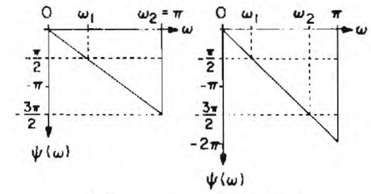
The equation $\operatorname{Re}(F(\omega)) = 0$ has the same roots as the equation

$$\phi_F(\omega) = \frac{\pi}{2} + k\pi, \text{ with } k \text{ an integer,}$$

since $F(\omega)$ has no zeros. From our previous discussion, we have

$$\begin{aligned}\phi_F(\pi) - \phi_F(0) &= \phi_B(\pi) - \phi_B(0) + \phi_D(\pi) - \phi_D(0) \\ &\leqslant -\left(\frac{N-1}{2}\right)\pi.\end{aligned}$$

Since the continuous function $\phi_F(\omega)$ decreases at least by $(N-1)/2\ \pi$ on the interval R, it follows that the graph of $\phi_F(\omega)$ crosses at least $N/2$ lines of phase $\pi/2 + k\pi$ in $(0, \pi]$ if N is even and at least $(N-1)/2$ such lines in $(0, \pi)$ if N is odd. Fig. 6 shows $\phi_F(\omega)$ when $b(n) = \delta(n)$, for the cases $N = 4$ and $N = 5$.



(a) (b)

Fig. 6. Unwrapped phase of the function $F(\omega)$ for $b(n) = \delta(n)$. (a) $N = 4$. (b) $N = 5$.

Statement A3: Let $a(n)$ be a real valued sequenced which is zero outside $0 \leqslant n \leqslant N-1$. If the equation

$$\operatorname{Im}\{A(z)\, z^{(N-1)/2}|_{z=e^{j\omega}}\} = 0$$

is satisfied for at least P distinct values of ω in the interval R, then it is identically equal to zero and $a(n) = a(N-1-n)$. P and R are defined as in (7) in the text as

$$P = \frac{N-1}{2} \text{ and } R = (0, \pi) \qquad \text{for } N \text{ odd}$$

$$P = \frac{N}{2} \text{ and } R = (0, \pi] \qquad \text{for } N \text{ even}$$

Proof for N Odd: With the use of trigonometric formulas, we obtain

$$\begin{aligned}G(\omega) &= \operatorname{Im}(A(\omega) e^{j\omega(N-1)/2}) \\ &= \sum_{n=0}^{N-1} a(n) \sin\left(\frac{N-1}{2} - n\right)\omega \end{aligned}\tag{A3.1}$$

$$G(\omega) = \sum_{n=1}^{(N-1)/2} \left\{a\left(\frac{N-1}{2} - n\right) - a\left(\frac{N-1}{2} + n\right)\right\} \sin n\omega. \tag{A3.2}$$

Since the set of the $(N-1)/2$ functions $\sin\omega, \sin 2\omega, \cdots, \sin(N-1)\omega/2$ is a Chebyshev set on the interval $(0, \pi)$ as is shown in [9] and since $G(\omega)$ has at least $(N-1)/2$ distinct roots in the interval $(0, \pi)$, it follows that the coefficients of the expansion in the right-hand side of (A3.2) must vanish

$$a\left(\frac{N-1}{2} - n\right) = a\left(\frac{N-1}{2} + n\right) = 0;$$

$$n = 1, 2, \cdots, \frac{N-1}{2}$$

or

$$a(n) = a(N-1-n); \qquad n = 0, 1, \cdots, N-1.$$

When N is even, the expansion of $G(\omega)$ is

$$G(\omega) = \sum_{n=0}^{(N/2)-1} \left\{a\left(\frac{N}{2} - 1 - n\right) - a\left(\frac{N}{2} + n\right)\right\} \cdot \sin\left(n + \frac{1}{2}\right)\omega.$$

Since the functions $\sin\omega/2, \sin 3\omega/2, \cdots, \sin N - 1/2\ \omega$ form a Chebyshev set on the interval $(0, \pi]$ as is shown in [9], it follows that

$$a\left(\frac{N}{2} - 1 - n\right) - a\left(\frac{N}{2} + n\right) = 0; \qquad n = 0, 1, \cdots, \frac{N}{2} - 1$$

or

$$a(n) = a(N - 1 - n); \qquad n = 0, 1, \cdots, N - 1.$$

This completes the proof of Statement A3.

References

[1] W. O. Saxton, *Computer Techniques for Image Processing in Electron Microscopy*. New York: Academic, 1978.

[2] G. N. Ramachandran and R. Srinivasan, *Fourier Methods in Crystallography*. New York: Wiley-Interscience, 1970.

[3] D. L. Misell, "An examination of an iterative method for the solution of the phase problem in optics and electron optics: I and II," *J. Phys. D. Appl. Phys.*, vol. 6, pp. 2200–2225, 1973.

[4] W. K. Pratt, *Digital Image Processing*. New York: Wiley, 1978.

[5] M. H. Hayes, "The reconstruction of a multidimensional sequence from the phase or magnitude of its fourier transform," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 140–154, Apr. 1982.

[6] M. H. Hayes, J. S. Lim, and A. V. Oppenheim, "Signal reconstruction from phase or magnitude," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 672–680, Dec. 1980.

[7] M. H. Hayes, "Signal reconstruction from phase or magnitude," Sc.D. dissertation, Dep. Elec. Eng. Comput. Sci., M.I.T., June 1981.

[8] Y. H. Bruck and L. G. Sodin, "On the ambiguity of the image reconstruction problem," *Opt. Commun.*, pp. 304–308, Sept. 1980.

[9] P. L. Van Hove, "Signal reconstruction from Fourier transform amplitude," S.M. thesis, Dep. Elec. Eng. Comput. Sci., M.I.T., Sept. 1982.

[10] R. M. Mersereau and D. E. Dudgeon, "The representation of two-dimensional sequences as one-dimensional sequences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-22, pp. 320–325, Oct. 1974.

[11] R. W. Gerchberg and W. O. Saxton, "A practical algorithm for the determination of phase from image and diffraction plane pictures," *Optik*, vol. 35, pp. 237–246, 1972.

[12] J. R. Fienup, "Space object imaging through the turbulent atmosphere," *Opt. Eng.*, vol. 18, pp. 529–534, Sept. 1979.

[13] A. V. Oppenheim, M. H. Hayes, and J. S. Lim, "Iterative procedure for signal reconstruction from Fourier transform phase," *Opt. Eng.*, vol. 21, pp. 122–127, Jan.–Feb. 1982.

Patrick L. Van Hove was born in Brussels, Belgium, on March 17, 1956. He received the Ingenieur Civil Electricien degree from the Universite Libre de Bruxelles (ULB), Brussels, Belgium, in 1978 and the M.S. in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1982.

From 1978 to 1981 he was an assistant at the ULB where he worked primarily on coherent optics and on acoustics. From September 1981 to August 1982, he was studying at M.I.T. under a fellowship from the Belgian American Educational Foundation, and he completed his thesis with the digital processing group. After completing his military duties in Belgium, he is now a Research Assistant with the Digital Signal Processing Group at M.I.T. His interests include theoretical aspects of signal processing, its application to image and speech processing, and optical holography.

Monson H. Hayes (S'76–M'80–S'80–M'81) was born in Washington, DC, on October 27, 1949. He received the B.S. degree in physics from the University of California, Berkeley, in 1971, and the S.M., E.E., and Sc.D. degrees, all in electrical engineering, from the Massachusetts Institute of Technology, Cambridge, in 1978, 1978, and 1981, respectively.

He worked as a Systems Engineer in IR systems technology at Aerojet Electrosystems from 1972 to 1974. From 1975 to 1979 he was a Teaching Assistant in the Department of Electrical Engineering at M.I.T., and from 1979 to 1981 was a Research Assistant at M.I.T. Lincoln Laboratory, Lexington, in the Multidimensional Digital Signal Processing Group. Since 1981 he has been with the Georgia Institute of Technology, Atlanta, where he is currently an Assistant Professor of Electrical Engineering. His research interests include digital signal processing and multidimensional signal processing and its applications to image processing.

Jae S. Lim (S'76–M'78) was born on December 2, 1950. He received the S.B., S.M., E.E., and Sc.D. degrees in electrical engineering and computer Science from the Massachusetts Institute of Technology Cambridge, in 1974, 1975, 1978, respectively.

He joined the M.I.T. faculty in 1978 as an Assistant Professor, and is currently Associate Professor in the Department of Electrical Engineering and Computer Science. His research interests include digital signal processing and its applications to image and speech processing. He has contributed more than 60 articles to journals and conference proceeedings, and is the editor of a reprint book, *Speech Enhancement* (Englewood Cliffs, NJ: Prentice-Hall, 1982). He is the winner of two prize paper awards, one from the Boston Chapter of the Acoustical Society of America in December 1976, and one from the IEEE Acoustics, Speech, and Signal Processing Society in April 1979.

Dr. Lim is a member of Eta Kappa Nu, Sigma Xi, and Chairman of the IEEE ASSP Society Technical Committee on Digital Signal Processing.

Alan V. Oppenheim (S'57–M'65–SM'71–F'77) received the S.B., S.M., and Sc.D. degrees from the Massachusetts Institute of Technology, Cambridge.

From 1961 to 1964 he was a member of the Massachusetts Institute of Technology Research Laboratory of Electronics and an Instructor in the Department of Electrical Engineering. During this period his research activities involved the application of modern algebra to the characterization of nonlinear systems and the development of homomorphic filtering as an approach to some nonlinear filtering problems. In 1964 he joined the faculty in the Department of Electrical Engineering and Computer Science, where he is currently full professor. He has been a Guggenheim Fellow at the University of Grenoble, Grenoble, France, he has held the Cecil H. Green Distinguished Chair in Electrical Engineering and Computer Science at M.I.T. and for a two year period he served as Associate Head of the Data Systems Division at M.I.T. Lincoln Laboratory. He has received a number of IEEE awards and the M.I.T. Graduate Student Council Teaching Award. His research interests are in the area of digital signal processing and its applications to speech, image, and seismic data processing, in the areas of system theory and knowledge-based signal processing. He is coauthor and editor of several texts in the area of signal processing, and is editor of a quarterly publication, *Trends & Perspectives in Signal Processing*.

Dr. Oppenheim is a member of Tau Beta Pi, Eta Kappa Nu, and Sigma Xi.